

ATES PATENT  
JUN 17 2004  
PATENT & TRADEMARK OFFICE

**Atty Docket No.      WILL.0003**

**Art Unit: 2186**

**REQUEST FOR PRIORITY  
UNDER 35 U.S.C. § 119  
AND THE INTERNATIONAL CONVENTION**

In the matter of the above-captioned application for a United States patent, notice is hereby given that the Applicant claims the priority date of December 25, 2003, the filing date of the corresponding Japanese Patent Application 2003-428718.

A certified copy of Japanese Patent Application 2003-428718 is being submitted herewith. Acknowledgment of receipt of the certified copy is respectfully requested in due course.

Respectfully submitted,

Stanley P. Fisher  
Registration Number 24,344

~~Juan Carlos A. Marquez~~  
~~Registration Number 34,072~~

**REED SMITH LLP**  
3110 Fairview Park Drive  
Suite 1400  
Falls Church, Virginia 22042  
(703) 641-4200  
**June 17, 2004**



日 本 国 特 許 庁  
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日                      2 0 0 3 年 1 2 月 2 5 日  
Date of Application:

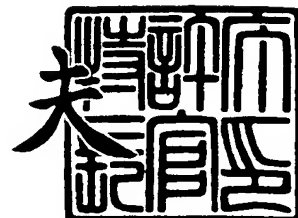
出 願 番 号                      特 願 2 0 0 3 - 4 2 8 7 1 8  
Application Number:  
[ST. 10/C] :                      [ J P 2 0 0 3 - 4 2 8 7 1 8 ]

出      願                      人                      株 式 会 社 日 立 製 作 所  
Applicant(s):

2 0 0 4 年    3 月 1 8 日

特許庁長官  
Commissioner,  
Japan Patent Office

今 井 康 夫



出証番号    出証特 2 0 0 4 - 3 0 2 2 0 0 2



【書類名】 特許願  
【整理番号】 340301447  
【あて先】 特許庁長官殿  
【国際特許分類】 G06F 07/24  
G06F 03/06

【発明者】  
【住所又は居所】 神奈川県小田原市中里 3 2 2 番 2 号 株式会社日立製作所 R A I  
D システム事業部内  
【氏名】 黒川 勇

【発明者】  
【住所又は居所】 神奈川県小田原市中里 3 2 2 番 2 号 株式会社日立製作所 R A I  
D システム事業部内  
【氏名】 武藤 淳一

【発明者】  
【住所又は居所】 神奈川県小田原市中里 3 2 2 番 2 号 株式会社日立製作所 R A I  
D システム事業部内  
【氏名】 竹内 久治

【発明者】  
【住所又は居所】 神奈川県小田原市中里 3 2 2 番 2 号 株式会社日立製作所 R A I  
D システム事業部内  
【氏名】 平松 伸一

【発明者】  
【住所又は居所】 神奈川県小田原市中里 3 2 2 番 2 号 株式会社日立製作所 R A I  
D システム事業部内  
【氏名】 塚田 大

【特許出願人】  
【識別番号】 000005108  
【氏名又は名称】 株式会社日立製作所

【代理人】  
【識別番号】 100095371  
【弁理士】  
【氏名又は名称】 上村 輝之

【選任した代理人】  
【識別番号】 100089277  
【弁理士】  
【氏名又は名称】 宮川 長夫

【選任した代理人】  
【識別番号】 100104891  
【弁理士】  
【氏名又は名称】 中村 猛

【手数料の表示】  
【予納台帳番号】 043557  
【納付金額】 21,000円

【提出物件の目録】  
【物件名】 特許請求の範囲 1  
【物件名】 明細書 1  
【物件名】 図面 1  
【物件名】 要約書 1  
【包括委任状番号】 0110323



**【書類名】 特許請求の範囲****【請求項 1】**

1 以上のホスト装置と通信チャネルを介して接続可能であり、前記ホスト装置による入出力の対象となるデータを記憶する 1 以上の物理デバイスを有する外部記憶装置と、前記ホスト装置と前記外部記憶装置との間における前記データの転送の制御を行うストレージ制御装置とを備えたストレージシステムにおいて、

前記ストレージ制御装置は、前記データの 1 以上のブロックを一時的に保持するキャッシュメモリと、制御情報を記憶する制御情報メモリと、前記ホスト装置と前記外部記憶装置と前記キャッシュメモリと前記制御情報メモリに接続され、前記外部記憶装置に記憶された前記データのブロックの 1 以上の集合をそれぞれ 1 以上の論理デバイスとして定義し、前記外部記憶装置と前記キャッシュメモリと前記ホスト装置の前記ブロックの入出力を制御する制御部とを有し、

前記制御部は、前記ホスト装置からソート処理実行指示を受けてソート処理を行う 1 以上のプロセッサを有し、

前記 1 以上のプロセッサは、任意の論理デバイス内の任意のブロック群をソートインブロック群として指定するソートイン情報と、前記ブロック内の任意の位置のデータをソートキーとして指定するキー情報と、任意の論理デバイス内の任意のブロック群をソートワークブロック群として指定するソートワーク情報と、任意の論理デバイス内の任意のブロック群をソートアウトブロック群として指定するソートアウト情報とを、前記ソート処理実行指示により通知され、

前記 1 以上のプロセッサは、前記ソート処理実行指示に応答して前記ソート処理を実行する場合、前記外部記憶装置内の前記指定されたソートインブロック群を前記キャッシュメモリに書き、前記指定されたソートワークブロック群を作業領域として使用しながら、前記指定されたソートキーの値に基づいて、前記キャッシュメモリに書かれた前記ソートインブロック群をソートし、ソートされた前記ソートインブロック群を前記指定されたソートアウトブロック群に格納し、そして、前記ソート処理の完了を前記ホスト装置に通知することを特徴とするストレージシステム。

**【請求項 2】**

請求項 1 記載のストレージシステムにおいて、

前記制御部に接続され、前記制御情報メモリの入出力情報を視覚的に表示できる管理機器を更に備え、

前記 1 以上のプロセッサは、前記ホスト装置からだけでなく前記管理機器からも前記ソート処理実行指示を受けて前記ソート処理を行えるように構成されていることを特徴とするストレージシステム。

**【請求項 3】**

請求項 1 記載のストレージシステムにおいて、

前記ソートイン情報は、前記ソートインブロック群が存在する論理デバイスのアドレスを示すソートインアドレスと、前記ソートインアドレスにより特定される前記論理デバイス内の前記ソートインブロック群が存在する場所範囲を示すソート範囲を含み、

前記ソートワーク情報は、前記ソートワークブロック群が存在する論理デバイスのアドレスを示すソートワークアドレスと、前記ソートワークアドレスにより特定される前記論理デバイス内の前記ソートワークブロック群が存在する場所範囲を示すソートワーク範囲を含み、

前記ソートアウト情報は、前記ソートアウトブロック群が存在する論理デバイスのアドレスを示すソートアウトアドレスと、前記ソートアウトアドレスにより特定される前記論理デバイス内の前記ソートアウトブロック群が存在する場所範囲を示すソートアウト範囲を含むことを特徴とするストレージシステム。

**【請求項 4】**

請求項 1 記載のストレージシステムにおいて、

前記 1 以上のプロセッサは、前記ソート処理実行指示により通知された前記ソートワー



ク情報が無効値をもつ場合、前記ソフトウェア情報により指定されるソフトウェアブロック群の代替の記憶領域を前記キャッシュメモリ内に確保し、そして、前記キャッシュメモリ内の前記代替の記憶領域を前記作業領域として使用することを特徴とするストレージシステム。

【請求項 5】

請求項 1 記載のストレージシステムにおいて、

前記制御部に接続され、前記制御情報メモリの入出力情報を視覚的に表示できる管理機器を更に備え、

前記管理機器は、任意の論理デバイス又は任意の論理デバイス内の任意のブロック群を、前記作業領域に使用されるために事前に確保される予約領域として指定する予約指示を前記制御部に出力することができるようになっており、

前記制御部の前記 1 以上のプロセッサは、前記管理機器からの前記予約指示に応答して、前記予約指示により指定された前記予約領域についてのアクセス禁止情報を前記制御情報メモリに格納し、前記制御情報メモリ内の前記アクセス禁止情報に基づいて、前記予約領域に対するホスト装置からのデータ入出力を制限し、そして、互いに重複する前記ソフトウェアブロック群を指定した複数の前記ソート処理実行指示を受けた場合、前記互いに重複するソフトウェアブロック群のいずれかの代替として前記予約領域を使用し、それにより、前記複数のソート処理実行指示を多重に実行するストレージシステム。

【請求項 6】

請求項 5 記載のストレージシステムにおいて、

前記 1 以上のプロセッサは、前記外部記憶装置内の前記 1 以上の物理デバイスのデータ入出力による使用状況をモニタして、前記 1 以上の物理デバイスの使用負荷量を示す物理デバイス使用情報を前記制御情報メモリに格納し、

前記 1 以上のプロセッサは、複数の前記予約領域が既に確保されている場合であって、前記複数の予約領域の中から前記代替として使用するものを選択する場合、前記制御情報メモリに記憶されている前記物理デバイス使用情報に基づいて、前記複数の予約領域の中から、前記使用負荷量が相対的に低い物理デバイスに存在する論理デバイスを選択することを特徴とするストレージシステム。

【請求項 7】

請求項 1 記載のストレージシステムにおいて、

前記論理デバイス内のブロックはそれぞれに固有の ID を有しており、

前記 1 以上のプロセッサは、前記外部記憶装置内の前記ソートインブロック群のソート処理において、前記ソートインブロック群内のソートされたブロックを、その ID を変更することなしに、前記ソフトウェアブロック群又は前記ソートアウトブロック群に書き、

前記 1 以上のプロセッサは、前記ソートインブロック群のソート処理の実行中又は完了後、前記ホスト装置のオンライン業務中に、前記外部記憶装置内の前記ソートインブロック群内のいずれかのブロックのデータを更新する場合、前記更新されるブロックについてオンラインソート処理を前記オンライン業務中に実行し、

前記 1 以上のプロセッサは、前記オンラインソート処理において、前記更新されるブロックを前記ソフトウェアブロック群又は前記ソートアウトブロック群内のソート後の位置に書くとともに、前記更新されるブロックの前記 ID に基づいて、前記ソフトウェアブロック群又は前記ソートアウトブロック群内から、前記更新されたブロックのソート前の位置に残存しているブロックを検出し削除することを特徴とするストレージシステム。

【請求項 8】

請求項 1 記載のストレージシステムにおいて、

前記 1 以上のプロセッサは、前記外部記憶装置内の前記ソートインブロック群のソート処理を行う場合、前記ソートインブロック群内の各ブロックの前記ストレージシステム内での位置と、前記ソートインブロック群内の各ブロックに対応する前記ソフトウェアブロック群又は前記ソートアウトブロック群内の各ブロックの前記ストレージシステム内での位置とを関連付けて表すブロック位置対応情報を前記制御情報メモリに記録し、



前記 1 以上のプロセッサは、前記ソートインブロック群のソート処理の実行中又は完了後、前記ホスト装置のオンライン業務中に、前記外部記憶装置内の前記ソートインブロック群内のいずれかのブロックのデータを更新する場合、前記更新されるブロックについてオンラインソート処理を前記オンライン業務中に実行し、

前記 1 以上のプロセッサは、前記オンラインソート処理において、前記更新されるブロックを前記ソートワークブロック群又は前記ソートアウトブロック群内のソート後の位置に書くとともに、前記ブロック位置対応情報に基づいて、前記ソートワークブロック群又は前記ソートアウトブロック群内から、前記更新されたブロックのソート前の位置に残存しているブロックを検出し削除し、さらに、前記更新されるブロックの前記ソートワークブロック群又は前記ソートアウトブロック群内での位置を前記ソート前の位置から前記ソート後の位置へと修正するように前記ブロック位置対応情報を更新することを特徴とするストレージシステム。

【請求項 9】

請求項 1 記載のストレージシステムにおいて、

前記 1 以上のプロセッサは、任意の論理デバイスの複製を作成し、前記複製の作成後、前記ホスト装置のオンライン業務中に、元の論理デバイス内のブロックを更新する場合、前記元の論理デバイス内の前記ブロックの更新を前記複製内のブロックに反映させる処理を前記オンライン業務中に行い、

前記 1 以上のプロセッサは、前記複製内のブロック群を前記ソートインブロック群とみなし、その後、前記オンライン業務中に、前記元の論理デバイス内の前記ブロックの更新を前記複製内のブロックに反映させる場合、前記複製内の前記更新が反映されるブロックについてのソート処理を前記オンライン業務中に実行することを特徴とするストレージシステム。

【請求項 10】

1 以上のホスト装置と通信チャネルを介して接続可能であり、前記ホスト装置による入出力の対象となるデータを記憶する外部記憶装置と、前記ホスト装置と前記外部記憶装置との間における前記データの転送の制御を行うストレージ制御装置とを備えたストレージシステムにおいて、

前記ストレージ制御装置は、所定の異なる用途に使用できる共用メモリと、前記ホスト装置と前記外部記憶装置と前記共用メモリに接続され、前記外部記憶装置と前記ホスト装置と前記共用メモリのデータ入出力を制御する制御部とを有し、

前記制御部は、前記ホスト装置からソート処理実行指示を受けてソート処理を行う 1 以上のプロセッサを有し、

前記 1 以上のプロセッサは、前記外部記憶装置内の第 1 の記憶領域をソートイン領域として選択し、前記外部記憶装置内又は前記共用メモリ内の第 2 の記憶領域をソートワーク領域として選択し、前記外部記憶装置内の第 3 の記憶領域をソートアウト領域として選択し、前記ソートワーク領域を用いて前記ソートイン領域のデータをソートし、ソートされたデータを前記ソートアウト領域に格納するというソート処理を行うようになっているストレージシステム。



**【書類名】明細書****【発明の名称】データソート機能を持つストレージシステム****【技術分野】****【0001】**

本発明は、一般には、ホスト装置と接続され、ホスト装置からの指示に応答してオンラインでデータを入出力するストレージシステムに関し、特に、ストレージシステムに蓄積されたデータをソートするための技術に関する。

**【背景技術】****【0002】**

ホスト装置と、そのホスト装置に対してオンラインでデータを入出力するストレージシステムと有する情報処理システムにおいて、ホスト装置によるストレージシステム内のデータへのアクセスの回数が増加することに伴い、ストレージシステム内のデータの配列の秩序が次第に乱れていく。そこで、随時にストレージシステム内のデータのソート処理を実行して、データを或るキー情報に従って整列させることは、ストレージシステムのデータ検索速度の低下を抑え、ホスト装置に対するデータ入出力の速度を高く維持するために重要である。例えば、銀行業務システムにおいては、顧客の口座のデータを、例えば口座番号又は顧客名などのキー情報に従って整列させるためのソート処理が、適当なタイミングで行われる必要がある。

**【0003】**

特許文献1に開示されたデータベース処理システムにおいては、ホスト処理装置に、データベース処理を行うデータ処理装置と、データを蓄積するディスク装置とが接続される。そして、データ処理装置が、データベース処理装置とソート処理装置を内蔵し、それにより、ソート処理とデータベース処理とを並行して行うことができる。ソート処理終了後に、ホスト処理装置が、ソート完了したデータを纏めてデータ処理装置から読み、これをディスク装置に対して出力する。

**【特許文献1】特開平8-272592号**

**【発明の開示】****【発明が解決しようとする課題】****【0004】**

今日、ストレージシステムの容量が増大し、行われるべきバックアップ処理やソート処理の数が増大している。とりわけ、従来技術に従うソート処理は、ホスト装置とストレージシステム間の頻繁な入出力処理を伴う。ソート処理数が増加すると、ホスト装置とストレージシステム間の入出力処理数が増加する。それにより、ホスト装置とストレージシステム間の通信経路の圧迫、ストレージシステム内のキャッシュメモリの使用効率の低下、及びストレージシステム内の同じ物理記憶デバイス（例えば、ハードディスクドライブ）へのアクセス競合によるタイムロスなどが生じ、入出力処理時間が増加する。結果として、ソート処理自体の時間の増加は勿論のこと、その影響を受けてバックアップ処理の要する時間も増大する。一般に、ソート処理やバックアップ処理は、オンライン業務に影響しないよう、オンライン業務が停止している夜間などに行われる。しかし、そうであっても、データのソート処理やバックアップ処理に要する時間が更に増加すれば、例えばオンライン業務の再開時刻が遅れるなど、情報処理システムの運用に重大な支障を与える可能性があることが危惧される。特に、24時間オンライン業務を行う場合には、オンラインランザクションのデータ入出力に影響することなしに、並行してソート処理が実行できる技術が必要である。

**【0005】**

従って、本発明の目的は、ストレージシステムにおけるソート処理の能力向上を図ることにある。

**【0006】**

本発明の別の目的は、ソート処理数が増加しても、ホスト装置とストレージシステム間の入出力処理数が増加しないようにする制御技術を提供することにある。



## 【0007】

本発明のまた別の目的は、ストレージシステム内のキャッシュメモリの使用効率を向上させる制御技術を提供することにある。

## 【0008】

本発明の更に別の目的は、ストレージシステム内の物理記憶デバイスへのアクセス競合を低減させる技術を提供することにある。

## 【0009】

本発明の更にまた別の目的は、オンライントランザクションのデータ入出力に大して影響することなしに、並行してソート処理が実行できる技術を提供することにある。

## 【課題を解決するための手段】

## 【0010】

本発明の一つの観点に従うストレージシステムは、1以上のホスト装置と通信チャネルを介して接続可能であり、前記ホスト装置による入出力の対象となるデータを記憶する1以上の物理デバイスを有する外部記憶装置と、前記ホスト装置と前記外部記憶装置との間における前記データの転送の制御を行うストレージ制御装置とを備える。前記ストレージ制御装置は、前記データの1以上のブロックを一時的に保持するキャッシュメモリと、制御情報を記憶する制御情報メモリと、前記ホスト装置と前記外部記憶装置と前記キャッシュメモリと前記制御情報メモリに接続され、前記外部記憶装置に記憶された前記データのブロックの1以上の集合をそれぞれ1以上の論理デバイスとして定義し、前記外部記憶装置と前記キャッシュメモリと前記ホスト装置の前記ブロックの入出力を制御する制御部とを有する。前記制御部は、前記ホスト装置からソート処理実行指示を受けてソート処理を行う1以上のプロセッサを有し、前記1以上のプロセッサは、任意の論理デバイス内の任意のブロック群をソートインブロック群として指定するソートイン情報と、前記ブロック内の任意の位置のデータをソートキーとして指定するキー情報と、任意の論理デバイス内の任意のブロック群をソートワークブロック群として指定するソートワーク情報と、任意の論理デバイス内の任意のブロック群をソートアウトブロック群として指定するソートアウト情報とを、前記ソート処理実行指示により通知される。そして、前記1以上のプロセッサは、前記ソート処理実行指示に応答して前記ソート処理を実行する場合、前記外部記憶装置内の前記指定されたソートインブロック群を前記キャッシュメモリに書き、前記指定されたソートワークブロック群を作業領域として使用しながら、前記指定されたソートキーの値に基づいて、前記キャッシュメモリに書かれた前記ソートインブロック群をソートし、ソートされた前記ソートインブロック群を前記指定されたソートアウトブロック群に格納し、そして、前記ソート処理の完了を前記ホスト装置に通知する。

## 【0011】

このストレージシステムによれば、ホスト装置がソート処理実行指示をストレージ制御装置に通知するだけで、ストレージ制御装置がソート処理実行指示に従い、指定されたブロック群を、指定されたキーの値に従ってソートし、そして、ソート後のブロックを指定されたブロック群に格納する。

## 【0012】

前記ストレージシステムの一つの実施形態は、前記制御部に接続され、前記制御情報メモリの入出力情報を視覚的に表示できる管理機器を更に備える。そして、前記1以上のプロセッサは、前記ホスト装置からだけでなく前記管理機器からも前記ソート処理実行指示を受けて前記ソート処理を行えるように構成されている。

## 【0013】

前記ストレージシステムの一つの実施形態においては、前記ソートイン情報は、前記ソートインブロック群が存在する論理デバイスのアドレスを示すソートインアドレスと、前記ソートインアドレスにより特定される前記論理デバイス内の前記ソートインブロック群が存在する場所範囲を示すソート範囲を含む。前記ソートワーク情報は、前記ソートワークブロック群が存在する論理デバイスのアドレスを示すソートワークアドレスと、前記ソートワークアドレスにより特定される前記論理デバイス内の前記ソートワークブロック群



が存在する場所範囲を示すソートワーク範囲を含む。また、前記ソートアウト情報は、前記ソートアウトブロック群が存在する論理デバイスのアドレスを示すソートアウトアドレスと、前記ソートアウトアドレスにより特定される前記論理デバイス内の前記ソートアウトブロック群が存在する場所範囲を示すソートアウト範囲を含む。

#### 【0014】

前記ストレージシステムの一つの実施形態においては、前記1以上のプロセッサは、前記ソート処理実行指示により通知された前記ソートワーク情報が無効値をもつ場合、前記ソートワーク情報により指定されるソートワークブロック群の代替の記憶領域を前記キャッシュメモリ内に確保し、そして、前記キャッシュメモリ内の前記代替の記憶領域を前記作業領域として使用する。

#### 【0015】

前記ストレージシステムの一つの実施形態は、前記制御部に接続され、前記制御情報メモリの入出力情報を視覚的に表示できる管理機器を更に備える。そして、前記管理機器は、任意の論理デバイス又は任意の論理デバイス内の任意のブロック群を、前記作業領域に使用されるために事前に確保される予約領域として指定する予約指示を前記制御部に出力することができるようになっている。また、前記1以上のプロセッサは、前記管理機器からの前記予約指示に応答して、前記予約指示により指定された前記予約領域についてのアクセス禁止情報を前記制御情報メモリに格納し、前記制御情報メモリ内の前記アクセス禁止情報に基づいて、前記予約領域に対するホスト装置からのデータ入出力を制限し、そして、互いに重複する前記ソートワークブロック群を指定した複数の前記ソート処理実行指示を受けた場合、前記互いに重複するソートワークブロック群のいずれかの代替として前記予約領域を使用する。これにより、複数のソート処理を多重に実行することができる。

#### 【0016】

この実施形態では、さらに、前記1以上のプロセッサは、前記外部記憶装置内の前記1以上の物理デバイスのデータ入出力による使用状況をモニタして、前記1以上の物理デバイスの使用負荷量を示す物理デバイス使用情報を前記制御情報メモリに格納する。そして、前記1以上のプロセッサは、複数の前記予約領域が既に確保されている場合であって、前記複数の予約領域の中から前記代替として使用するものを選択する場合、前記制御情報メモリに記憶されている前記物理デバイス使用情報に基づいて、前記複数の予約領域の中から、前記使用負荷量が相対的に低い物理デバイスに存在する論理デバイスを選択する。これにより、外部記憶装置の物理デバイス使用負荷量を均一化することができ、入出力処理能力のバランス化を図ることができる。

#### 【0017】

前記ストレージシステムの一つの実施形態では、前記論理デバイス内のブロックはそれぞれに固有のIDを有している。前記1以上のプロセッサは、前記外部記憶装置内の前記ソートインブロック群のソート処理において、前記ソートインブロック群内のソートされたブロックを、そのIDを変更することなしに、前記ソートワークブロック群又は前記ソートアウトブロック群に書くようになっている。そして、前記1以上のプロセッサは、前記ソートインブロック群のソート処理の実行中又は完了後、前記ホスト装置のオンライン業務中に、前記外部記憶装置内の前記ソートインブロック群内のいずれかのブロックのデータを更新する場合、前記更新されるブロックについてオンラインソート処理を前記オンライン業務中に実行する。このオンラインソート処理において、前記1以上のプロセッサは、前記更新されるブロックを前記ソートワークブロック群又は前記ソートアウトブロック群内のソート後の位置に書くとともに、前記更新されるブロックの前記IDに基づいて、前記ソートワークブロック群又は前記ソートアウトブロック群内から、前記更新されたブロックのソート前の位置に残存しているブロックを検出し削除する。

#### 【0018】

前記ストレージシステムの一つの実施形態では、前記1以上のプロセッサは、前記外部記憶装置内の前記ソートインブロック群のソート処理を行う場合、前記ソートインブロッ



ク群内の各ブロックの前記ストレージシステム内での位置と、前記ソートインブロック群内の各ブロックに対応する前記ソートワークブロック群又は前記ソートアウトブロック群内の各ブロックの前記ストレージシステム内での位置とを関連付けて表すブロック位置対応情報を前記制御情報メモリに記録する。そして、前記1以上のプロセッサは、前記ソートインブロック群のソート処理の実行中又は完了後、前記ホスト装置のオンライン業務中に、前記外部記憶装置内の前記ソートインブロック群内のいずれかのブロックのデータを更新する場合、前記更新されるブロックについてオンラインソート処理を前記オンライン業務中に実行する。このオンラインソート処理においてし、前記1以上のプロセッサは、前記更新されるブロックを前記ソートワークブロック群又は前記ソートアウトブロック群内のソート後の位置に書くとともに、前記ブロック位置対応情報に基づいて、前記ソートワークブロック群又は前記ソートアウトブロック群内から、前記更新されたブロックのソート前の位置に残存しているブロックを検出し削除し、さらに、前記更新されるブロックの前記ソートワークブロック群又は前記ソートアウトブロック群内での位置を前記ソート前の位置から前記ソート後の位置へと修正するように前記ブロック位置対応情報を更新する。

#### 【0019】

前記ストレージシステムの一つの実施形態では、前記1以上のプロセッサは、任意の論理デバイスの複製を作成し、前記複製の作成後、前記ホスト装置のオンライン業務中に、元の論理デバイス内のブロックを更新する場合、前記元の論理デバイス内の前記ブロックの更新を前記複製内のブロックに反映させる処理を前記オンライン業務中に行うようになっている。さらに、前記1以上のプロセッサは、前記複製内のブロック群を前記ソートインブロック群とみなし、その後、前記オンライン業務中に、前記元の論理デバイス内の前記ブロックの更新を前記複製内のブロックに反映させる場合、前記複製内の前記更新が反映されるブロックについてのソート処理を前記オンライン業務中に実行する。

#### 【0020】

本発明の別の観点に従うストレージシステムは、1以上のホスト装置と通信チャネルを介して接続可能であり、前記ホスト装置による入出力の対象となるデータを記憶する外部記憶装置と、前記ホスト装置と前記外部記憶装置との間における前記データの転送の制御を行うストレージ制御装置とを備える。前記ストレージ制御装置は、所定の異なる用途に使用できる共用メモリと、前記ホスト装置と前記外部記憶装置と前記共用メモリに接続され、前記外部記憶装置と前記ホスト装置と前記共用メモリのデータ入出力を制御する制御部とを有する。前記制御部は、前記ホスト装置からソート処理実行指示を受けてソート処理を行う1以上のプロセッサを有し、前記1以上のプロセッサは、前記外部記憶装置内の第1の記憶領域をソートイン領域として選択し、前記外部記憶装置内又は前記共用メモリ内の第2の記憶領域をソートワーク領域として選択し、前記外部記憶装置内の第3の記憶領域をソートアウト領域として選択し、前記ソートワーク領域を用いて前記ソートイン領域のデータをソートし、ソートされたデータを前記ソートアウト領域に格納するというソート処理を行うようになっている。

#### 【0021】

本発明の好適な実施形態によれば、ストレージシステムにおけるソート処理の能力向上を図ることができる。

#### 【0022】

本発明の好適な実施形態によれば、ソート処理数が増加しても、ホスト装置とストレージシステム間の入出力処理数が大きく増加しないようにすることができる。

#### 【0023】

本発明の好適な実施形態によれば、ストレージシステム内のキャッシュメモリの使用効率を向上させることができる。

#### 【0024】

本発明の好適な実施形態によれば、ストレージシステム内の物理記憶デバイスへのアクセス競合を低減させることができる。



## 【0025】

本発明の好適な実施形態によれば、オンライントランザクションのデータ入出力に大して影響することなしに、並行してソート処理が実行できる。

## 【発明を実施するための最良の形態】

## 【0026】

本発明の実施形態に係る情報処理システムについて、図面を参照しながら詳細に説明する。

## 【0027】

図1は、本発明の一実施形態にかかるストレージシステムを用いた情報処理システムの全体構成を示す。

## 【0028】

図1に示す情報処理システムは、例えば業務情報処理のような情報処理を行うホスト装置1と、ホスト装置1に接続され、ホスト装置1により利用されるデータを蓄積し、そしてホスト装置1からの指示に応じてホスト装置1に対してデータの入出力を行うストレージシステム2とを備える。図1の例では、1つのホスト装置1がストレージシステム2に接続されているが、複数のホスト装置1がストレージシステム2に接続されてもよい。ホスト装置1とストレージシステム2は、両者間でのデータやデータアクセス指示などの入出力の制御を行うチャンネル3を介して、通信可能に接続される。ストレージシステム2は、外部記憶装置11とストレージ制御装置12を備える。外部記憶装置11は、データを蓄積するための1又は複数（典型的には多数）の物理的な記憶デバイス（例えばハードディスクドライブであり、以下、「物理デバイス」という）（図示省略）を有する。ストレージ制御装置11は、ホスト装置1からのデータアクセス指示やその他各種の指示の処理、外部記憶装置13とホスト装置1の間でのデータ転送の制御、及び本発明の原理に従う外部記憶装置11内のデータのソート処理などを行う機能をもつ。

## 【0029】

ストレージ制御装置12は、後述する各種の制御機能をもつ制御部121、制御部121によって使用される各種の制御用の情報を記憶した制御情報メモリ122、外部記憶装置11及びホスト装置1に対して入出力されるデータを一時的に記憶するキャッシュメモリ123を有する。制御部121は、チャンネル3、制御情報メモリ122、キャッシュメモリ123及び外部記憶装置11に通信可能に接続される。制御部121は、管理機器14（ストレージシステム2の状態監視や保守点検などの目的に使用される、例えば、パーソナルコンピュータ）にも接続される。制御部121は、ホスト装置1からデータアクセス指示（リード指示、ライト指示など）を受信し解釈しホスト装置1へ然るべき応答を返すこと、ホスト装置1から受け取ったライトデータをキャッシュメモリ123に書くこと、キャッシュメモリ123に書かれたライトデータを外部記憶装置11に転送すること、ホスト装置1から指示されたリードデータを外部記憶装置11から取得してキャッシュメモリ123に書くこと、キャッシュメモリ123に書かれたリードデータをキャッシュメモリ123からホスト装置1に送ること、などのホスト装置1に対するデータ入出力処理に関わる制御を行う機能をもつ。さらに、制御部121は、ホスト装置1からデータソート指示を受信し解釈しホスト装置1へ然るべき応答を返すこと、ホスト装置1から受け取ったデータソート指示に従って後に詳述するようにしてデータのソート処理を行うこと、などのデータソート処理に関わる制御を行う機能をもつ。制御部121は、ホスト装置1からのデータソート指示だけでなく、管理機器14からのデータソート指示に応答してデータソート制御を行うこともできる。制御部121は、前記の機能以外に、データの保護やバックアップに関わる機能や他の様々な機能をもつが、これらの機能についてはこの明細書では説明を省略する。制御部121は、1又は複数（典型的には複数）のプログラムされたマイクロプロセッサ（図示省略）を有し、それらのマイクロプロセッサが上述した各種の制御を分担して処理する。

## 【0030】

制御情報メモリ122には、制御部21が上述したデータ入出力処理に関わる制御やデ



ータソート処理に関わる制御やその他の制御を行うときに、制御部 21 によって使用される（例えば、読まれ、書かれ、又は変更される）ことになる各種の制御情報が記憶される。とりわけ、データソート処理に関わる制御において使用される制御情報については、後に詳述する。

#### 【0031】

制御部 121 は、外部記憶装置 11 の物理的な記憶領域を論理的な記憶領域に変換し、また、論理的な記憶領域を物理的な記憶領域に変換する機能をもつ。制御部 121 は、この物理／論理記憶領域変換機能を用いて、外部記憶装置 11 の記憶領域に 1 又は複数（典型的には多数）の論理的な記憶デバイス（以下、「論理デバイス」という）112 を定義する。そして、制御部 121 は、ホスト装置 1 が所望の論理デバイス 112 に対してデータ入出力指示を発することができるように、論理デバイス 112 をベースにした通信インタフェースをホスト装置 1 に提供する。

#### 【0032】

管理機器 14 は、制御情報メモリ 122 に対する入出力機能と、制御情報メモリ 122 に対する入出力情報を視覚的に表示できる機能などを持つ。

#### 【0033】

図 2 は、外部記憶装置 11 の構成、とりわけ物理デバイス 111 と論理デバイス 121 との関係及び論理デバイス 121 の詳細な構成を示す。

#### 【0034】

図 2 に示すように、外部記憶装置 11 は、1 又は複数（典型的には多数）の物理デバイス 131 を有する。上述した制御部 121 によって、それら物理デバイス 131 が提供する記憶領域上に、1 又は複数の（典型的には多数）の論理デバイス 112 が定義される。論理デバイス 112 は、1 つの物理デバイスの記憶領域を制御上の論理範囲で分割した形態で定義されることができる。或いは、論理デバイス 112 は、グループ化された 2 つ以上の物理デバイス 111 の記憶領域を制御上の論理範囲で分割した形態で、例えば、RAID の原理に従って 2 つ以上の物理デバイス 111 に分散された領域をから構成されるようにて、定義されることができる。ホスト装置 1 が入出力する対象となるデータは、制御部 121 によって、論理デバイス 112 毎に管理される。

#### 【0035】

一般に、1 つの論理デバイス 112 は、制御上のより小さい論理範囲に分割され、分割された個々の領域は論理トラック 20 として定義される。1 つの論理トラック 20 は、制御上の更に細かい論理範囲（典型的には最少単位）で分割されて、分割された個々の領域はレコード 21 として定義される。ホスト装置 1 がストレージ制御装置 12 に対して送信するデータの入出力指示は、入出力対象のデータが存在する論理デバイス 112 のアドレスと、論理トラック 20 の番号（トラック番号）とレコード 21 の番号（レコード番号）を指定することで、その入出力対象のデータを特定する。ホスト装置 1 からの入出力指示の対象となったデータは、通常、論理トラック 20 毎に、レコード 21 という単位でキャッシュメモリ 123 及び物理デバイス 111 に記憶される。

#### 【0036】

外部記憶装置 13 内の全ての論理デバイス 20、トラック 20、及びレコード 21 を定義した情報は、制御部 121 により、制御情報メモリ 122 に記録される。また、制御部 121 は、外部記憶装置 13 内のそれぞれの物理デバイス 111 の使用状況のモニタリングを行い、その結果得られた物理デバイス 111 の使用状況の情報を制御情報メモリ 122 に記録する。

#### 【0037】

以下、この実施形態におけるデータのソート処理について詳細に説明する。尚、以下の説明では、理解の容易のために、外部記憶装置 13 内の単一の物理デバイス内に定義された論理デバイスを例にとる。1 つの論理デバイスがグループ化された複数の物理デバイスに跨るように定義される、いわゆる RAID 構成、が採用される場合には、その論理デバイスの使用状況は、その論理デバイスを構成する各物理デバイスの使用状況から判明する



。よって、以下の単一の物理デバイスを例に取ったソート処理の説明から、RAID構成におけるソート処理も、当業者は容易に理解可能である。また、本発明は、データのソート自体に必要なソートアルゴリズムを限定するものではなく、情報処理技術の中で一般に使用される様々なソートアルゴリズムが適用できる。そのソートアルゴリズムを行うためのソートプログラムは、例えば図1に示した制御部121内のマイクロプロセッサにプログラムされることができる。

【0038】

まず、ホスト装置1がストレージシステム2のストレージ制御装置12にソート処理を指示する動作に関して説明する。

【0039】

ソート処理を実行するソートプログラムがホスト装置1にインストールされている。利用者は、ホスト装置1上で、ソートプログラムを起動して例えばバッチ業務としてのソートジョブの開始をソートプログラムに指示するためのソートジョブデータを、所定のジョブ制御言語（JCL）を用いて作成する。このソートジョブデータの作成において、利用者は、ソート処理の実行に必要な複数種類のパラメータのセットを定義する。図3には、利用者によって定義されることができるパラメータセット30の種類が例示されている。図3に示すように、例えば次のような種類のパラメータ301～309のセット30がソートジョブデータに含まれる。

【0040】

(1) ソートINアドレス

ソートINアドレスとは、ソート対象となる論理デバイスのアドレスである。

【0041】

(2) ソート範囲

ソート範囲とは、ソートINアドレスによって指定された論理デバイス内でのソート対象となるブロック群（データセット）が存在する範囲である。ソート範囲は、例えば開始トラック番号と終了トラック番号によって指定される。

【0042】

(3) ソートキー

ソートキーとは、ソートの基準として使用されるキー（項目）の値である。

【0043】

(4) キー位置

キー位置とは、ブロック内でソートキーが存在する位置である。

【0044】

(5) ソートOUTアドレス

ソートOUTアドレスとは、ソートされ終わったデータブロック群が記憶される論理デバイスのアドレスである。

【0045】

(6) ソートOUT範囲

ソートOUT範囲とは、ソートOUTアドレスによって指定された論理デバイス内でソートされ終わったブロック群（ソートOUTデータセット）が記憶される範囲である。ソートOUT範囲は、例えば開始トラック番号と終了トラック番号によって指定される。

【0046】

(7) ソートWORKアドレス

ソートWORKアドレスとは、ソート処理の作業領域として使用される論理デバイスのアドレスである。

【0047】

(8) ソートWORK範囲

ソートWORK範囲とは、ソートWORKアドレスによって指定された論理デバイス内で作業領域として使用されるブロック群（ソートWORKデータセット）が記憶される範囲である。ソートOUT範囲は、例えば開始トラック番号と終了トラック番号によって指



定される。

【0048】

(9) ソート方向

ソート方向とは、ソートされたデータの配列順序の方向であり、これには例えば、昇順また降順などが指定される。

【0049】

ホスト装置1では、上記のようなパラメータセット30が定義されると、そのパラメータセット30を含んだソートジョブデータがエントリされ、そして、ホスト装置1のOSが、このソートジョブデータを用いてソートプログラムを起動する。ソートジョブデータを受けて、ホスト装置1のソートプログラムは、ストレージシステム2のストレージ制御装置12に対してソート処理の実行を指示する。ストレージ制御装置12は、その指示に応じてソート処理を行う。

【0050】

図5は、ホスト装置1がストレージ制御装置12にソート処理の実行を指示し、そして、ストレージ制御装置12にソート処理を実行する動作の流れを示す。

【0051】

図5において、ステップ501では、ホスト装置1のソートプログラムが、前述のJCLによるソートジョブデータを加工して、図4に例示するようなパラメータセット40を生成する。このパラメータセット40には、図3に示した利用者により定義されたパラメータセット30と同じ内容のパラメータが含まれている。そして、ホスト装置1のソートプログラムは、ソート処理実行指示のコマンドをストレージシステム2のストレージ制御装置12に対して発行する。ステップ502では、ストレージ制御装置12が、ソート処理実行指示のコマンドを受けて、ホスト装置1からパラメータセット40を受領する。ステップ503では、ホスト装置1が、ソートジョブを待ち状態を登録し、ストレージ制御装置12からのソートジョブ終了通知を待つ。但し、ホスト装置1は、ジョブ終了待ち状態を登録するだけで、待っている間、オンライン業務などの別の処理を実行することができる。

【0052】

ステップ504では、ストレージ制御装置12が、パラメータセット40を解析し、指定されたソートINアドレス401、ソート範囲402、ソートOUTアドレス403、ソートOUT範囲404、ソートWORKアドレス405、ソートWORK範囲406、キー407、キー位置408及びソート方向などのパラメータの内容を把握する。そして、ストレージ制御装置12内の制御部121が、把握されたパラメータの内容に従ってソート処理を開始する。

【0053】

ソート処理においては、ステップ505では、外部記憶装置13内の指定されたソートINアドレス401内のソート範囲402に存在する複数のレコード21が、キャッシュメモリ123へ転送される。ステップ506では、制御部121が、ソートINアドレス401のソート範囲402からキャッシュメモリ123に転送されたソート対象の複数のレコード21を、指定されたキー407とキー位置408に従いソートする。ステップ507では、制御部121が、現在のソート処理123上に必要の無いレコード21を、キャッシュメモリ123から指定されたソートWORKアドレス405内のソートWORK範囲406へ書き出すことで退避させる。ステップ508では、先にソートWORKアドレス405のソートWORK範囲406へ退避されたが、現在のソート処理に必要となったレコード21が、ソートWORK範囲406から再度キャッシュメモリ123へ読み出される。

【0054】

ステップ509では、制御部121が、ソート対象の全てのレコード21のソートが完了したかを判断し、ソートが未完了であれば、制御をステップ506へ戻し、ソートされていないレコード21について同様のソートを行う。他方、ステップ509でソートが完



了したと判断された場合には、制御はステップ510へ進む。ステップ510では、ソートの完了したキャッシュメモリ123上のレコード21及びソートWORKアドレス405のソートWORK範囲406に退避されているレコード21が、開始トラック番号から終了トラック番号に向かうトラック番号順で順次に、ソートOUTアドレス403のソートOUT範囲404へ転送される。ソートされた全てのレコードのソートOUT範囲404への格納が完了後、制御はステップ511へ進み、ストレージ制御装置12が、ソートジョブが終了したことをホスト装置1へ通知し、ホスト装置1は、ソートジョブ終了待ち状態の登録を消去する。

#### 【0055】

上記説明から分かるように、ソート処理が完了するまでは、ソートされているレコード21の位置は、ソートOUTアドレス403とソートOUT範囲404又はソートWORKアドレス405とソートWORK範囲406内の位置で示されることができる。

#### 【0056】

ホスト装置1からだけでなく、管理機器14からの指示によりストレージ制御装置12にソート処理を行わせるようにすることもできる。その場合の制御は、図3～図5を参照して行った上の説明においてホスト装置1を管理機器14と読み替えることで説明され得る。

#### 【0057】

上記の制御により、ソート処理が、ホスト装置1及びチャンネル3を使用せずにストレージシステム2の内部で行われる。そのため、ソート処理数が増加しても、ホスト装置1とストレージシステム2間の入出力処理数が増加しない。

#### 【0058】

さて、上述の説明では、ストレージシステム2内でソート処理の作業領域として使用される論理デバイス112のアドレス（ソートWORKアドレス405）とブロック範囲（ソートWORK範囲406）が、ホスト装置1又は管理機器14などの外部機器によって指定される。しかし、ホスト装置1か又は管理機器14ら指定されるソートWORKアドレス405とソートWORK範囲406が無効である（例えば、実在しない、或いは、現在別の目的で使用されているためワーク作業領域に割り当てられない、など）場合が発生し得る。このような場合、ストレージシステム2のストレージ制御装置12は、自動的にソート処理用の作業領域を確保することができる。すなわち、このような場合、ストレージ制御装置12は、例えばキャッシュメモリ123上の未使用領域をソート処理用の作業領域として確保し、これをソート処理終了まで作業領域として占有することを、自動的に実施する。

#### 【0059】

図6は、ストレージ制御装置12が自動的にソート処理用の作業領域を確保する処理の流れを示す。

#### 【0060】

図6に示す制御は、ホスト装置1又は管理機器14などの外部機器から指定されたソートWORKアドレス405とソートWORK範囲406が無効である場合に、ストレージ制御装置12の制御部121によって開始されるステップ601では、制御部121が、外部機器から指定されたソートINアドレス401のソート範囲402の持つデータ量より大きいサイズのキャッシュメモリ123上の未使用領域を、ソート処理の作業領域として確保し、そして、ソートWORK範囲406を、キャッシュメモリ123上で確保された作業領域の位置情報へ置換する。そして、制御部121はソート処理を開始する。

#### 【0061】

ソート処理において、ステップ602では、外部記憶装置13内の指定されたソートINアドレス401のソート範囲402に存在するレコード21が、キャッシュメモリ123のソートWORK範囲406（作業領域）へ読み込まれる。ステップ603では、制御部121が、キャッシュメモリ123のソートWORK範囲406に読み込まれたレコード21を、指定されたキー407、キー位置408及びソート方法409に従ってソート



する。ステップ604では、現在のソート処理に必要な無くなったキャッシュメモリ123の作業領域上のレコード21が、そのキャッシュメモリ123のソートWORK範囲406内の未使用部分へ退避される。ステップ605では、先にキャッシュメモリ123のソートWORK範囲406内の未使用部分へ退避されたが、現在のソートに必要なレコード21が再度復帰されて、ソートWORK範囲406上の現在のソート対象のレコード21へ組み込まれる。ステップ606では、ソートが完了したかが判断され、ソートが未完了であれば、制御はステップ603へ戻る。他方、ソートが完了した場合には、制御はステップ607へ進む。ステップ607では、ソートされたキャッシュメモリ123上のソートWORK範囲406上のレコード21が、ソートOUTアドレス403のソートOUT範囲404へ、開始トラック番号から終了トラック番号までトラック番号の順で順次に、転送される。このとき、キャッシュメモリ123上のソートWORK範囲406上に存在する、ソートINアドレス401のソート範囲402が持つデータ量分のレコードが、ソートOUT範囲404へ転送される。出力完了後、制御はステップ608へ進み、ソートジョブ終了通知がホスト装置1又は管理機器14へ送られ、ホスト装置1又は管理機器14は、ソートジョブ終了待ち状態の登録を消去する。

#### 【0062】

上記の制御により、ソート処理の作業領域としてキャッシュメモリ123が自動的に利用され、キャッシュメモリ123の使用効率が向上する。また、ストレージシステム内の物理デバイス111へのアクセス競合が低減する。

#### 【0063】

さて、ストレージ制御装置12は、特定の論理デバイス112の特定のブロック範囲を、ソート処理の作業領域に使用可能な領域に、予め割り当てる機能をもつこともできる。ストレージ制御装置12は、そのような作業領域の事前割当のための指示を、例えば管理機器14から受けることができる。

#### 【0064】

作業領域として事前に割当てられた論理デバイス112のブロック範囲については、ホスト装置1からのアクセスが禁止される必要がある。このアクセス禁止を制御するために、図7に例示するように、制御情報メモリ122内に、各論理デバイス112の状態を表した論理デバイス管理テーブル70が記憶され、論理デバイス管理テーブル70に、論理デバイス112毎に、ホスト装置1によるアクセスが禁止されているか否かを示すアクセス禁止フラグ701を設定することができる。図7に示された論理デバイス管理テーブル70の例では、それぞれの論理デバイス112の論理デバイスアドレスに対応して管理情報が記録されており、この管理情報には、アクセス禁止フラグ701と、その論理デバイス112が存在する物理デバイス111の場所を示す情報（例えば、物理デバイス111の物理デバイスアドレス及び物理デバイス111内のブロック範囲（トラック番号範囲）など）が含まれている。そして、アクセス禁止フラグ701の値「1」が、ホスト装置1によるアクセスが禁止されていることを意味し、値「0」はそのアクセスが禁止されていないことを意味する。

#### 【0065】

管理機器14は、任意の論理デバイス112をソート処理の作業領域として選択する際、その選択された論理デバイス112の論理デバイスアドレスとアクセス禁止フラグ701の値「1」をストレージ制御装置12に通知する。ストレージ制御装置12は、制御情報メモリ122内の論理デバイス管理テーブル70上の、通知された論理デバイスアドレスに対応するアクセス禁止フラグ701に、通知された値「1」を設定する。ストレージ制御装置12は、論理デバイス管理テーブル70を参照し、アクセス禁止フラグ701に値「1」が設定されている論理デバイス112については、ホスト装置1からの使用を禁止する。これにより、ソート処理の作業領域として使用可能なソートWORKアドレス405とソートWORK範囲406が、予め1つ以上確保される。この状態においては、ストレージ制御装置12は、1又は複数のホスト装置1から、ソートWORKアドレス405とソートWORK範囲406が重複する複数のソート処理実行指示を受領した場合、い



ずれかのソート処理実行指示により指定されたソートWORKアドレス405とソートWORK範囲406の代替として、予め作業領域として確保しておいた論理デバイス112内の記憶領域を利用することができる。この場合、ストレージ制御装置12は、指定されたソートWORKアドレス405とソートWORK範囲406を、その代替として利用される記憶領域の論理デバイスアドレスとブロック範囲に置き換える。このようにして、多重にソート処理を実行することができる。

#### 【0066】

予め作業領域のために確保しておいた1以上の論理デバイス112の中から、上述のような代替の作業領域を選択する場合に、ストレージ制御装置12は、次のように動作することができる。すなわち、ストレージ制御装置12は、図7に示すような制御情報メモリ122内の論理デバイス管理テーブル70及び物理デバイス管理テーブル71を参照する。物理デバイス管理テーブル71には、それぞれの物理デバイス111のアドレスである物理デバイスに対応して、それぞれの物理デバイス111に対する過去のデータ入出力指示の回数（つまり、それぞれの物理デバイス111の使用負荷量）711が含まれている。因みに、この入出力指示回数711は、ストレージ制御装置12が、継続的に、受領したデータ入出力指示を物理デバイス111毎にカウントして記録しているものである。さて、ストレージ制御装置12は、論理デバイス管理テーブル70及び物理デバイス管理テーブル71に基づいて、作業領域に使用可能なように予め確保されている論理デバイス112の記憶領域の中から、ソート処理実施要求により指定されたソートINアドレス401及びソートOUTアドレス403で示される論理デバイス112が存在する物理デバイス131とは異なる物理デバイス111上に存在する記憶領域だけを、選択候補として絞り込む。そして、ストレージ制御装置12は、絞り込まれたそれぞれの記憶領域が存在している物理デバイス111の入出力指示回数711を参照して、入出力指示回数711が最も少ない物理デバイス111に存在する記憶領域を、代替の作業領域として選択する。そして、ストレージ制御装置12は、その選択された代替の作業領域の論理デバイスアドレスとブロック範囲を、図4に示されたパラメータセットのソートWORKアドレス405とソートWORK範囲406に設定する。このようにして、過去の使用負荷が相対的に低い物理デバイス131上に存在する記憶領域を優先的に作業領域に割り当てることにより、記憶領域による使用負荷量の偏りを減少させることができる。また、物理デバイス111へのアクセス競合が低減する。

#### 【0067】

上述したソート処理は、ホスト装置1のオンライン業務が行われていない期間に行うことができることは言うまでも無い。しかし、オンライン業務処理中にこれと並行してソート処理を実行することもできる。以下に、オンライン業務処理中にソート処理を実行する場合の動作を説明する。

#### 【0068】

図2に示すように、ストレージ制御装置12は、或る論理デバイス112内にレコード21を新規作成する際、各レコード21に対してその論理デバイス112内で固有なID214を割当て、各レコード21にそのレコードID214を記録する。すなわち、ストレージ制御装置12は、ホスト装置1からのデータアクセス指示を受けた場合、その指示がレコード新規作成指示か否かを判断する。判断結果が真ならば、ストレージ制御装置12は、ホスト装置1から転送されてくるレコード21のデータの後端に固有ID214を付加する。他方、判断結果が偽であるならば、ストレージ制御装置12は、指示通りのアクセスを実施するが、その際、固有ID214については何の操作も行わない。従って、各レコード21の固有ID214は、レコード21が更新された場合には、変化しない。ストレージ制御装置12は、固有ID214を含めたレコード21に対してソート処理を実行し、ソート処理において固有ID214を変化させない。よって、ストレージ制御装置12は、ソート処理完了後、所望のレコード21の固有ID214をソートOUTアドレス403のソートOUT範囲404から検索することで、そのレコード21のソート後の位置を知ることができる。



## 【0069】

更に、ストレージ制御装置12は、図8中に示すようなソートレコード位置対応テーブル80を制御情報メモリ122に設定する。ソートレコード位置対応テーブル80には、ソートIN側情報81とソートWORK/OUT側情報82が相互に関連付けられて記憶される。ソートIN側情報81には、ソートINアドレス401のソート範囲402内に存在する全てのレコード21（つまり、ソート対象の全てのレコード21）の、ソートINアドレス401内での論理トラック番号211とレコード番号212が記録される。ソートWORK/OUT側情報82には、ソート対象の全てのレコード21のソートOUTアドレス403又はソートWORKアドレス405内での論理トラック番号211とレコード番号212が記録される（ソート処理中、ソートOUTアドレス403に存在するレコード21についてはソートOUTアドレス403内の番号、ソートOUTアドレス403には存在せずソートWORKアドレス405に存在するレコード21についてはソートWORKアドレス405内の番号）。ストレージ制御装置12は、ソートレコード位置対応テーブル80を使用することにより、ソートINアドレス401のソート範囲402内の論理トラック番号211とレコード番号212で一意に決まるレコード21が、ソートOUTアドレス403のソートOUT範囲404内の論理トラック番号211とレコード番号212における、または、ソートWORKアドレス405のソートWORK範囲406内の論理トラック番号211とレコード番号212におけるどのレコード21に対応するかが認知できる。

## 【0070】

図9は、ストレージ制御装置12がソートレコード位置対応テーブル80を作成する処理の流れを示す。

## 【0071】

図9において、ステップ901では、ストレージ制御装置12がソート対象のレコード21の全てをソートIN側情報81に登録する。登録される内容は、ソートINアドレス401、ソートINアドレス401内のトラック番号211、及びトラック20内レコード番号212である。ステップ902では、ソート処理が実行される。ソート処理中又はソート処理後に、ステップ903で、ストレージ制御装置12が、各レコード21のソート先位置を決定し、ソートWORK/OUT側情報82に登録する。登録される内容は、ソートWORKアドレス405若しくはソートOUTアドレス403、ソートWORKアドレス405若しくはソートOUTアドレス403内のトラック番号211、及びトラック20内のレコード番号212である。ステップ904で、ストレージ制御装置12がソート処理完了かをチェックし、未完であればステップ902へ戻る。こうして、ソート処理中もしくは完了後に、ソートレコード位置対応テーブル80が作成される。

## 【0072】

図10は、ストレージ制御装置12が、オンライン業務中にバックグラウンドでソート処理を行う動作例の概要を示す。

## 【0073】

図10に示すように、ストレージ制御装置12が、任意の論理デバイス1000の複製1001を作成する複製機能と、複製1001の作成後に元の論理デバイス1000内のレコード21が更新された場合にその更新されたレコード21のみを複製1001に反映するスナップショット機能を有する。ストレージ制御装置12は、ソート対象の論理デバイス（以下、「ソートIN」という）1000の複製1001を、上述したソート処理におけるソートINアドレスにより指定された論理デバイスとして扱う。オンライン業務中、ストレージ制御装置12は、ソートIN1000内のいずれかのレコード21が更新されると、そのレコード21の更新をその複製1001に反映させる。さらに、ストレージ制御装置12は、複製1001のレコード21の更新処理に同期して、複製1001内の更新されたレコード21のソートを実施して、ソートされたレコード21を別の論理デバイス（以下、「ソートOUT」という）1002に格納する。

図11は、ストレージ制御装置12がオンライン業務中にバックグラウンドでソート処



理を行う動作の流れを示す。

【0074】

図11において、ステップ1101では、ストレージ制御装置12が、ソートINとして指定された論理アドレス1000の複製1001を、ソートINと認識し、この複製1001のアドレスをソートINアドレス401に設定する。ステップ1102では、前述のソートレコード位置対応テーブル80の設定が行われる。すなわち、ストレージ制御装置12が、ソート対象のレコード（複製1001内のレコード）21の全てをソートレコード位置対応テーブル80のソートIN側情報81に登録する。登録される内容は、ソートINアドレス（複製1001のアドレス）401、複製1001内トラック番号211、及びトラック211内レコード番号212である。ステップ1103では、ソート処理が実行される。ソート処理中又は完了後に、ステップ1104では、ストレージ制御装置12が、ソート先のレコード21の位置を決定し、ソートレコード位置対応テーブル80のソートWORK/OUT側情報82に登録する。登録される内容は、ソートWORKアドレス405又はソートOUTアドレス403、ソートWORKアドレス405又はソートOUTアドレス403内トラック番号211、及びトラック211内レコード番号213である。ステップ1105でソート処理完了かがチェックされ、未完であれば制御はステップ1103へ戻る。こうして、ソート処理中もしくは完了後に、ソートレコード位置対応テーブル80が作成される。

【0075】

ステップ1106では、ストレージ制御装置12が、ソート処理が完了した後も常にソート再開可能な状態になる。

【0076】

ステップ1107では、ストレージ制御装置12が、複製側1001、すなわち、ソートINアドレスでのレコード更新が発生したかをチェックする。更新が発生した場合は、制御はステップ1108へ進み、ストレージ制御装置12が、更新されたソートINアドレス内のトラック番号211のレコード番号212のレコード21のみを対象として、ソート処理を実施する。この更新されたレコード21のみを対象としたソート処理では、その対象のレコード21が、ソートWORKアドレス又はソートOUTアドレス内のソート後の新たな位置に書かれる。それとともに、その対象のレコード21のIDを用いて、ソートWORKアドレス又はソートOUTアドレス内のソート前の位置に残存している対象のレコード21が検出されて消去される。こうして、更新されたレコード21が、他のレコード21と正しい順序で整列するように、ソートWORKアドレス又はソートOUTアドレスにマージされる。

【0077】

ステップ1109では、上記の更新されたレコード21のみを対象としたソート処理中又は処理後に、ストレージ制御装置12が、今回マージされたレコード21について、ソートレコード位置対応テーブル80のソートWORK/OUT側情報82に記録されたレコード位置（ソートWORKアドレス又はソートOUTアドレス内のトラック番号211及びレコード番号212）を更新する。

【0078】

ステップ1110では、ストレージ制御装置12が、今回マージされたレコード21のWORK/OUT側情報82内の更新されたレコード位置211、212が、更新前より小さいかチェックする。チェックの結果が肯定の場合には、制御はステップ1111へ進む。ステップ1111では、ストレージ制御装置12が、ソート処理の一環として、ソートWORKアドレス又はソートOUTアドレスにおいて、今回マージされたレコード21の一つ後のレコード位置から、今回マージされたレコード21のソート前のソートWORK/OUT側情報82が示していたレコード位置より一つ前のレコード位置までの範囲に位置する全てのレコード21のレコード番号212を-1する方向にずらして、ソートWORKアドレス又はソートOUTアドレスに書き直す。それに伴い、ステップ1115では、ストレージ制御装置12が、ステップ1114でレコード番号212が更新されたレ



コード21について、ソートレコード位置対応テーブル80のソートWORK/OUT側情報82を更新する。

【0079】

また、ステップ1113では、ストレージ制御装置12が、今回マージされたレコード21のソートWORK/OUT側情報82内の更新されたレコード位置211、212が、更新前より大きいチェックする。ソート結果が肯定の場合は、制御はステップ1114へ進む。ステップ1114では、ストレージ制御装置12が、ソート処理の一環として、ソートWORKアドレス又はソートOUTアドレスにおいて、今回マージされたレコード21の一つ前のレコード位置から、今回マージされたレコード21のソート前のソートWORK/OUT側情報82が示していたレコード位置の一つ後のレコード位置までの範囲に位置する全てのレコード21のレコード番号212を-1する方向にずらして、ソートWORKアドレス又はソートOUTアドレスに書き直す。それに伴い、ステップ1115では、ストレージ制御装置12が、ステップ1114でレコード番号212が更新されたレコード21について、ソートレコード位置対応テーブル80のソートWORK/OUT側情報82を更新する。

【0080】

以上のようにして、オンライン業務中は、更新されたレコードのみに限定してソート処理が行われる。これにより、オンライン業務に大した支障を与えずに、オンライン業務と並行してソート処理ができる。なお、図11中のステップ1108～1105で示すオンライン業務中の更新データのソート処理は、ステップ1101～1104で示す複製のソート処理の完了後だけでなく、複製のソート処理の実行中にブロック更新が発生した時に行うこともできる。

【0081】

既に説明したように、ソート処理はホスト装置1からの指令によるだけでなく、管理機器14からの指令によっても行えるようにしてもよい。管理機器14が図4に示したようなパラメータセット40をストレージ制御装置12へ通知することで、ストレージ制御装置12は、そのパラメータセット40に従ってソート処理を行うことができる。管理機器14からの指示によるソート処理が完了した場合、ストレージ制御装置12はソート処理完了を、管理機器14だけでなくホスト装置1にも報告することができる。

【0082】

また、上述した予めソート処理の作業領域のための論理デバイスを確保する処理は、管理機器14からの指示によるだけでなく、ホスト装置1からの指示によって行えるようにすることもできる。いずれからの指令によって論理デバイスを確保する処理を行った場合でも、ストレージ制御装置12は、どの論理デバイスが確保されたかを、管理機器14及びホスト装置1に報告することができる。

【0083】

本発明が適用されるストレージシステム2の構成は図1に示したようなものである必要は無く、他の構成のストレージシステムにも本発明が適用できる。図12は、本発明が適用できるストレージシステムの別の構成の一例を示している。図12の構成例では、複数のホスト装置1がストレージ制御装置12に接続されている。ストレージ制御装置12は、制御部121と、これに接続された複数の共有メモリ125を備える。制御部121は、複数のチャネルアダプタ1211と、複数のディスクアダプタ1212を有する。各チャネルアダプタ1211は、ホスト装置1及び全ての共有メモリ125に接続され、ホスト装置1との通信及びホスト装置1と送受するデータや指令の処理を制御を行う。各ディスクアダプタ1212は、全ての共有メモリ125及び外部記憶装置11内の全ての物理デバイス111に接続され、物理デバイス111との通信及び物理デバイス111と送受するデータや指令の処理を制御を行う。共有メモリ125には、図1に示した制御情報メモリ122と同様の用途に使用される制御情報メモリ領域1251と、図1に示したキャッシュメモリ123と同様の用途に使用されるキャッシュメモリ領域1252とがある。このような構成のストレージシステムにおいても、ストレージ制御装置12が既に説明し



たような制御を行うことにより、ホスト装置 1 のオンライン業務のデータ入出力に大した影響を与えずにソート処理を行うことができる。

【0084】

上述した実施形態では、ストレージシステム 2 は図 1 に示したキャッシュメモリとして使用されるような構成をもつが、他の構成のストレージシステムにも本発明が適用できることは言うまでもない。例えば、図 12 は、ストレージシステムの別の構成例を示しているが、これにも本発明は適用できる。

【0085】

以上、本発明の実施形態を説明したが、これは本発明の説明のための例示であり、本発明の範囲を上記実施形態に限定する趣旨ではない。本発明は上記実施形態以外の様々な態様で実施することができる。

【図面の簡単な説明】

【0086】

【図 1】 本発明の一実施形態にかかるストレージシステムを用いた情報処理システムの全体構成を示す図である。

【図 2】 外部記憶装置 11 の構成、とりわけ物理デバイス 111 と論理デバイス 121 との関係及び論理デバイス 121 の詳細な構成を示す図である。

【図 3】 利用者が定義するソート処理に必要なパラメータセット 30 を例示した図である。

【図 4】 図 3 のパラメータセット 30 に基づいてホスト装置が作成するパラメータセット 40 を例示した図である。

【図 5】 ホスト装置 1 がストレージ制御装置 12 にソート処理の実行を指示し、そして、ストレージ制御装置 12 にソート処理を実行する制御の流れを示す図である。

【図 6】 ストレージ制御装置 12 が自動的にソート処理用の作業領域を確保する制御の流れを示す図である。

【図 7】 制御情報メモリ 122 内の論理デバイス管理テーブル 70 及び物理デバイス管理テーブル 71 の構成を示す図である。

【図 8】 制御情報メモリ 122 内のソートレコード位置対応テーブル 80 の構成を示す図である。

【図 9】 ストレージ制御装置 12 がソートレコード位置対応テーブル 80 を作成する処理の流れを示す図である。

【図 10】 ストレージ制御装置 12 がオンライン業務中にソート処理を行う動作の概要を示す図である。

【図 11】 ストレージ制御装置 12 がオンライン業務中にソート処理を行う動作の流れを示す図である。

【図 12】 本発明が適用できるストレージシステムの別の構成の一例を示す図である。

【符号の説明】

【0087】

- 1・・・ホスト処理装置
- 2・・・ストレージシステム
- 11・・・外部記憶装置
- 12・・・ストレージ制御装置
- 121・・・制御部
- 122・・・制御情報メモリ
- 123・・・キャッシュメモリ
- 111・・・物理記憶デバイス
- 112・・・論理記憶デバイス
- 20・・・論理トラック
- 21・・・レコード



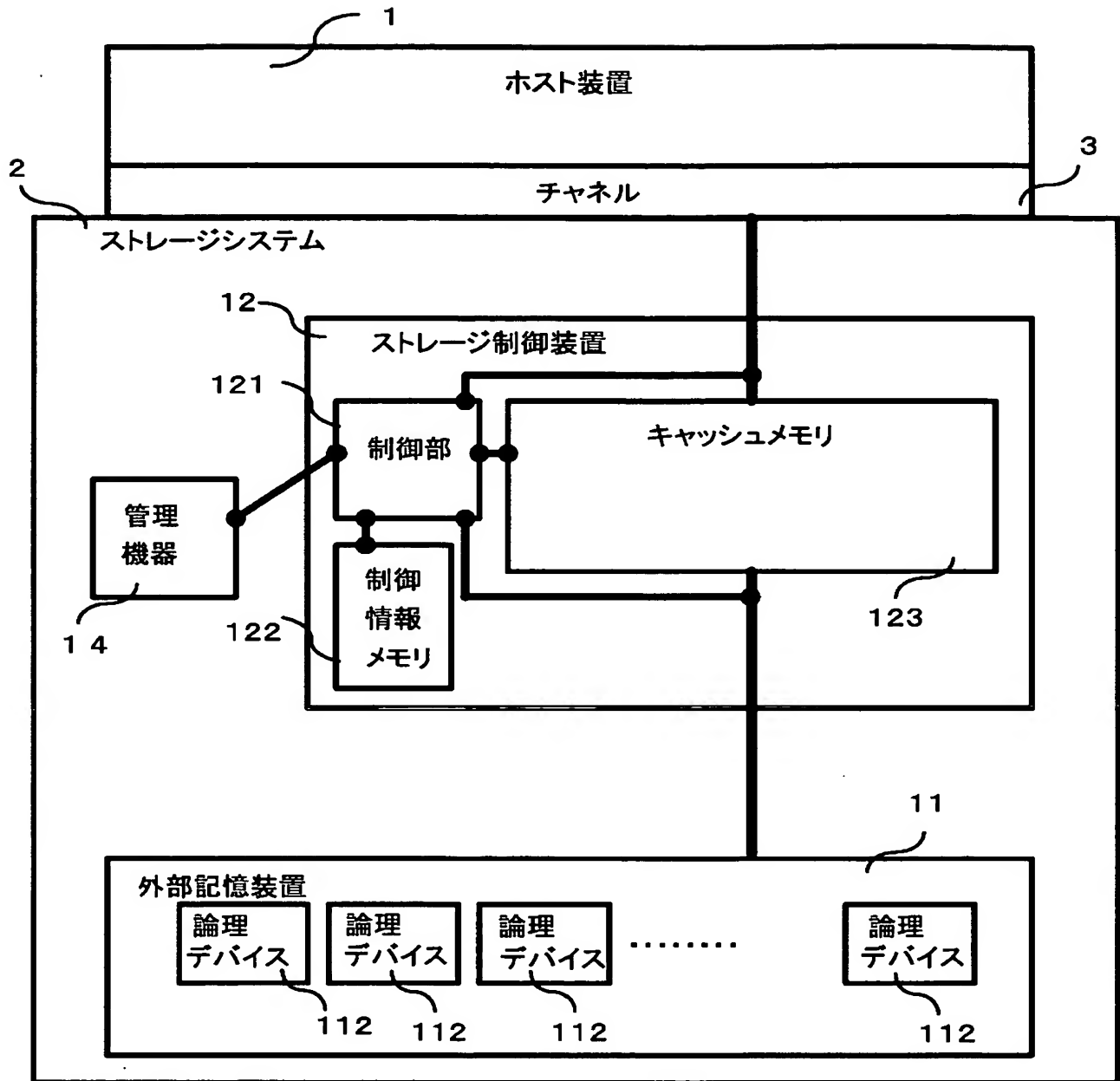
2 1 1 . . . ト ラ ッ ク 番 号

2 1 2 . . . レ コ ー ド 番 号

2 1 4 . . . 固 有 I D

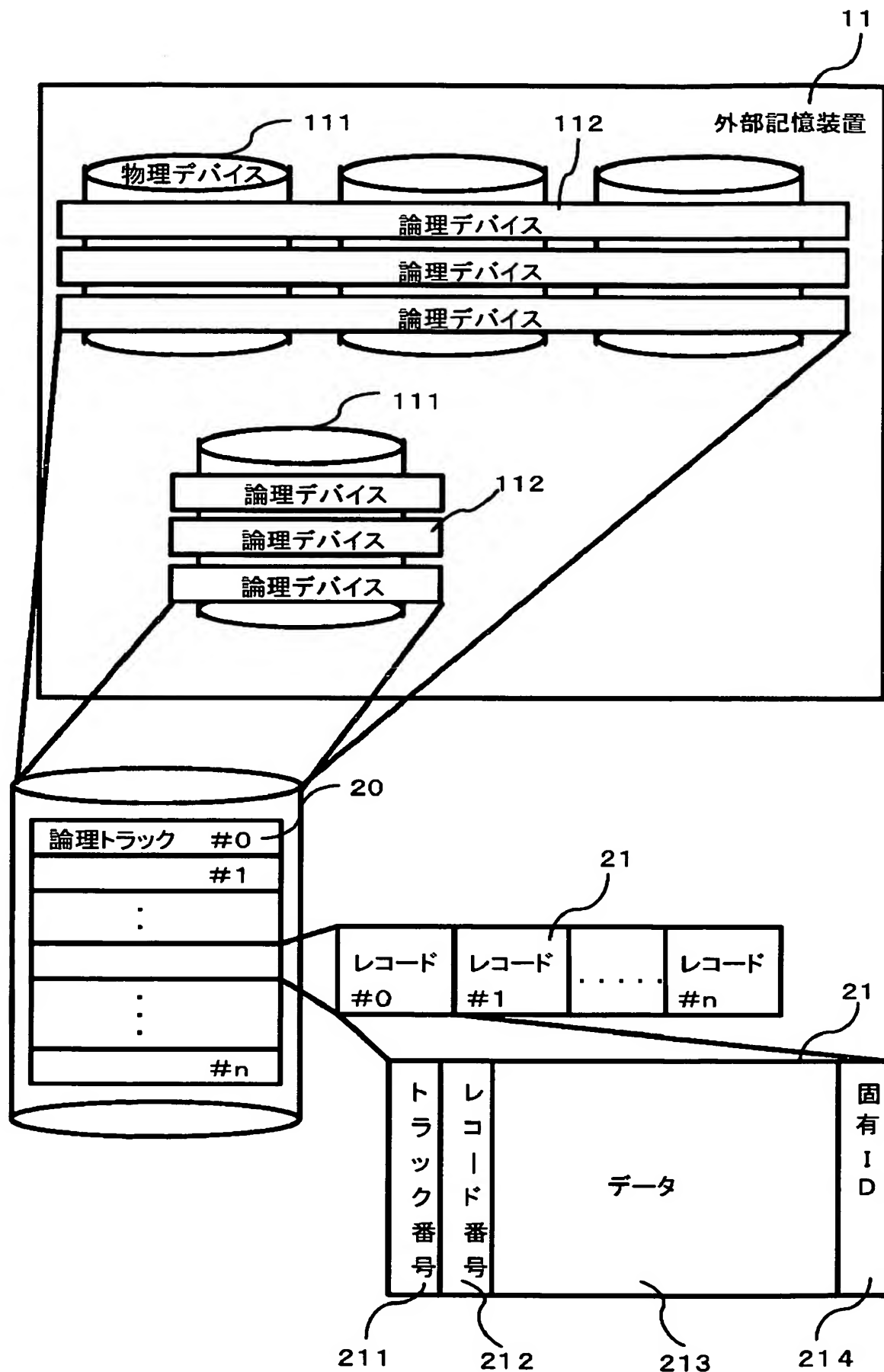


【書類名】図面  
【図1】





【図 2】





【図 3】

30

	パラメータ種別	内容	
1	ソートINアドレス	ソート対象となる論理デバイスのアドレス	301
2	ソート範囲 (開始トラック番号、 終了トラック番号)	上記論理デバイス内でソート対象となるブロック群が存在する範囲 (ソート対象データセット)	302
3	ソートキー	ソートの基準となるキーの値	303
4	キー位置	ブロック内でソートキーが存在する位置	304
5	ソートOUTアドレス	ソートされ終わったデータブロック群が記憶される論理デバイスのアドレス	305
6	ソートOUT範囲 (開始トラック番号、 終了トラック番号)	上記論理デバイス内でソートされ終わったブロック群が記憶される範囲 (ソートOUTデータセット)	306
7	ソートWORKアドレス	ソート処理の作業領域として使用される論理デバイスのアドレス	307
8	ソートWORK範囲 (開始トラック番号、 終了トラック番号)	上記論理デバイスで作業領域として使用されるブロック群の範囲 (ソートWORKデータセット)	308
9	ソート方向	ソートの順序の方向 (昇順、降順、etc. )	309

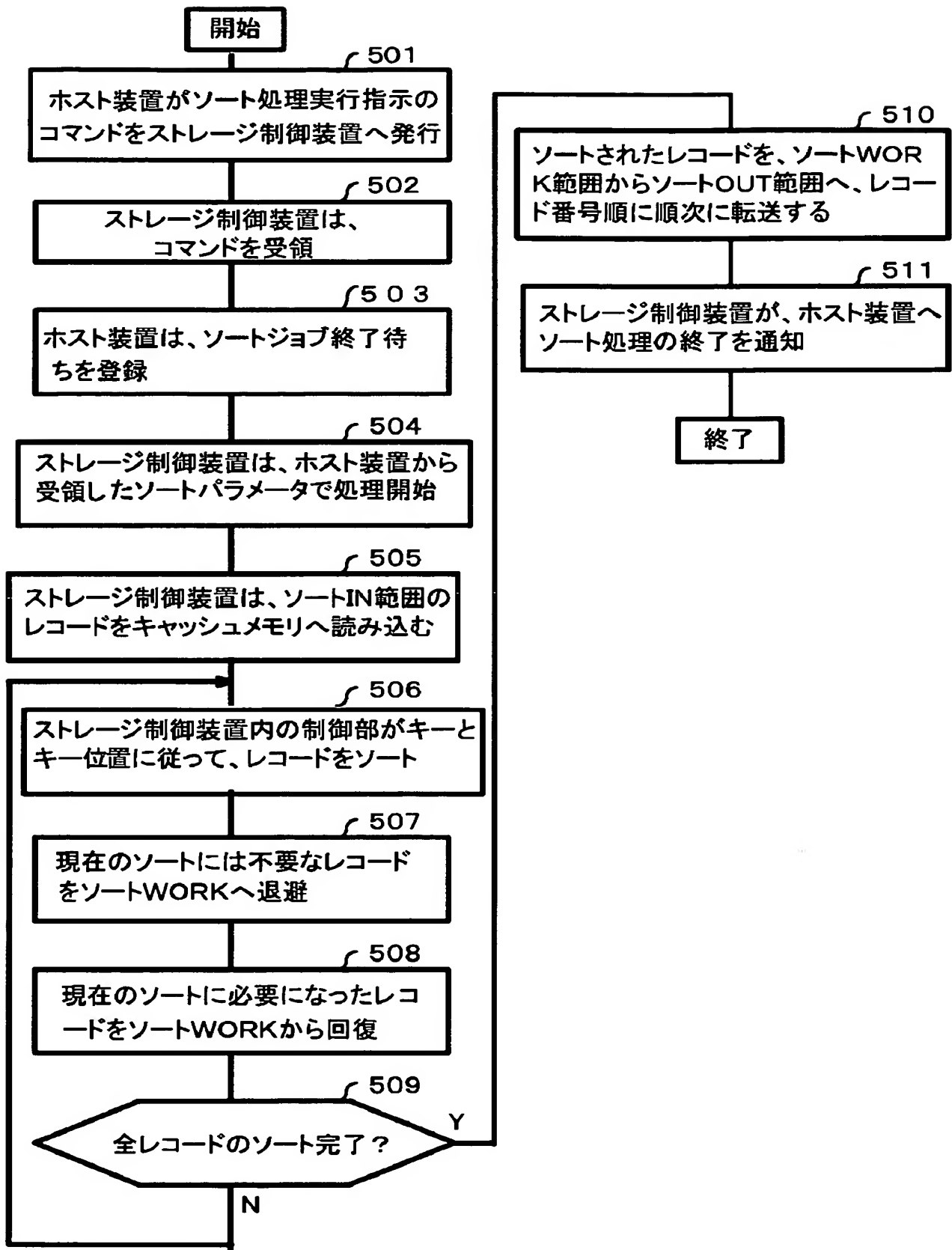


【図 4】

40		
	パラメータ種別	内容
1	ソートINアドレス	ソート対象となる論理デバイスのアドレス
2	ソート範囲 (開始トラック番号、 終了トラック番号)	上記論理デバイス内でソート対象となるブロック群が存在する範囲 (ソート対象データセット)
3	ソートOUTアドレス	ソートされ終わったデータブロック群が記憶される論理デバイスのアドレス
4	ソートOUT範囲 (開始トラック番号、 終了トラック番号)	上記論理デバイス内でソートされ終わったブロック群が記憶される範囲 (ソートOUTデータセット)
5	ソートWORKアドレス	ソート処理の作業領域として使用される論理デバイスのアドレス
6	ソートWORK範囲 (開始トラック番号、 終了トラック番号)	上記論理デバイスで作業領域として使用されるブロック群の範囲 (ソートWORKデータセット)
7	ソートキー	ソートの基準となるキーの値
8	キー位置	データブロック内でソートキーが存在する位置
9	ソート方向	ソートの順序の方向 (昇順、降順、etc.)

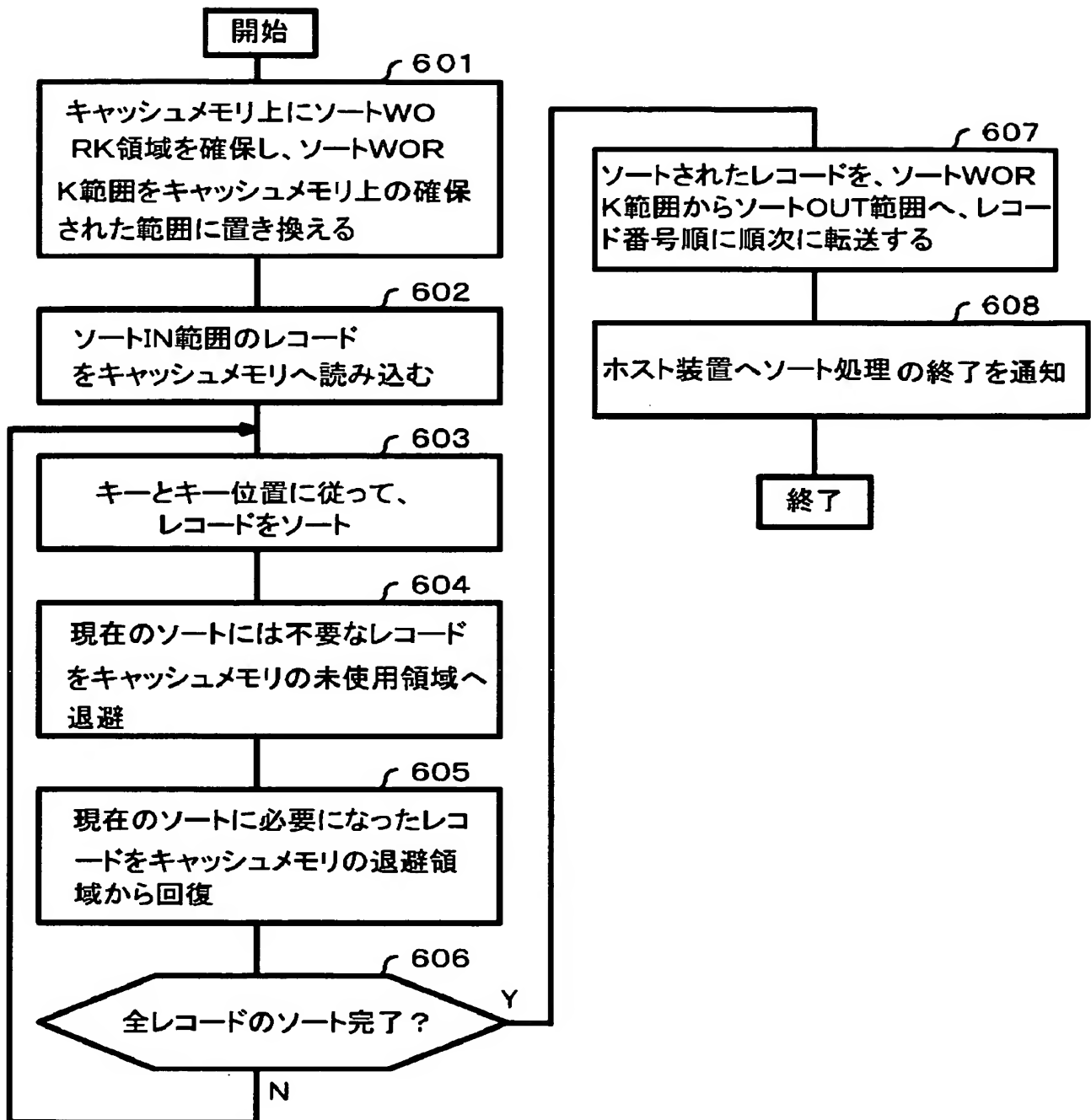


【図 5】



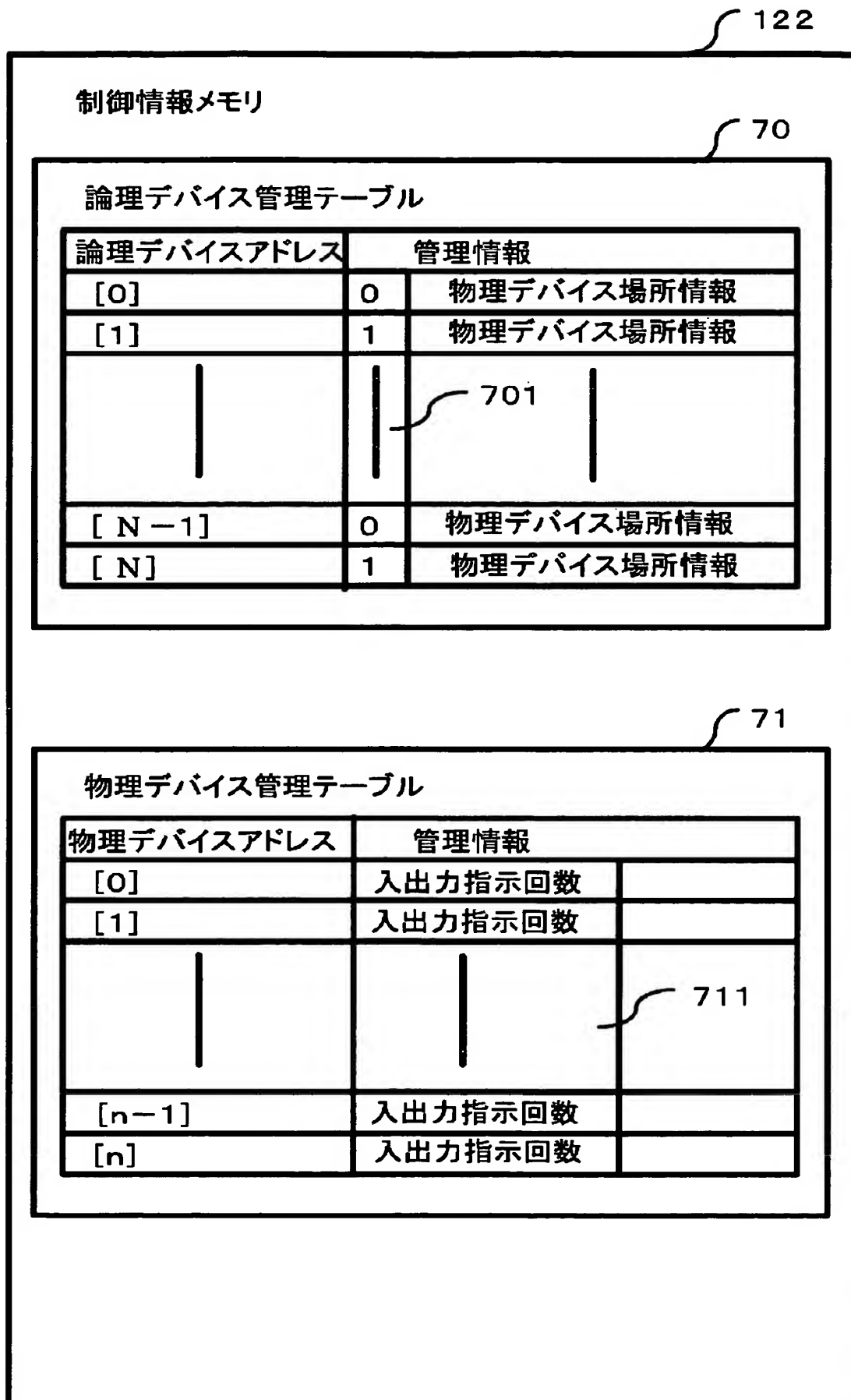


【図 6】



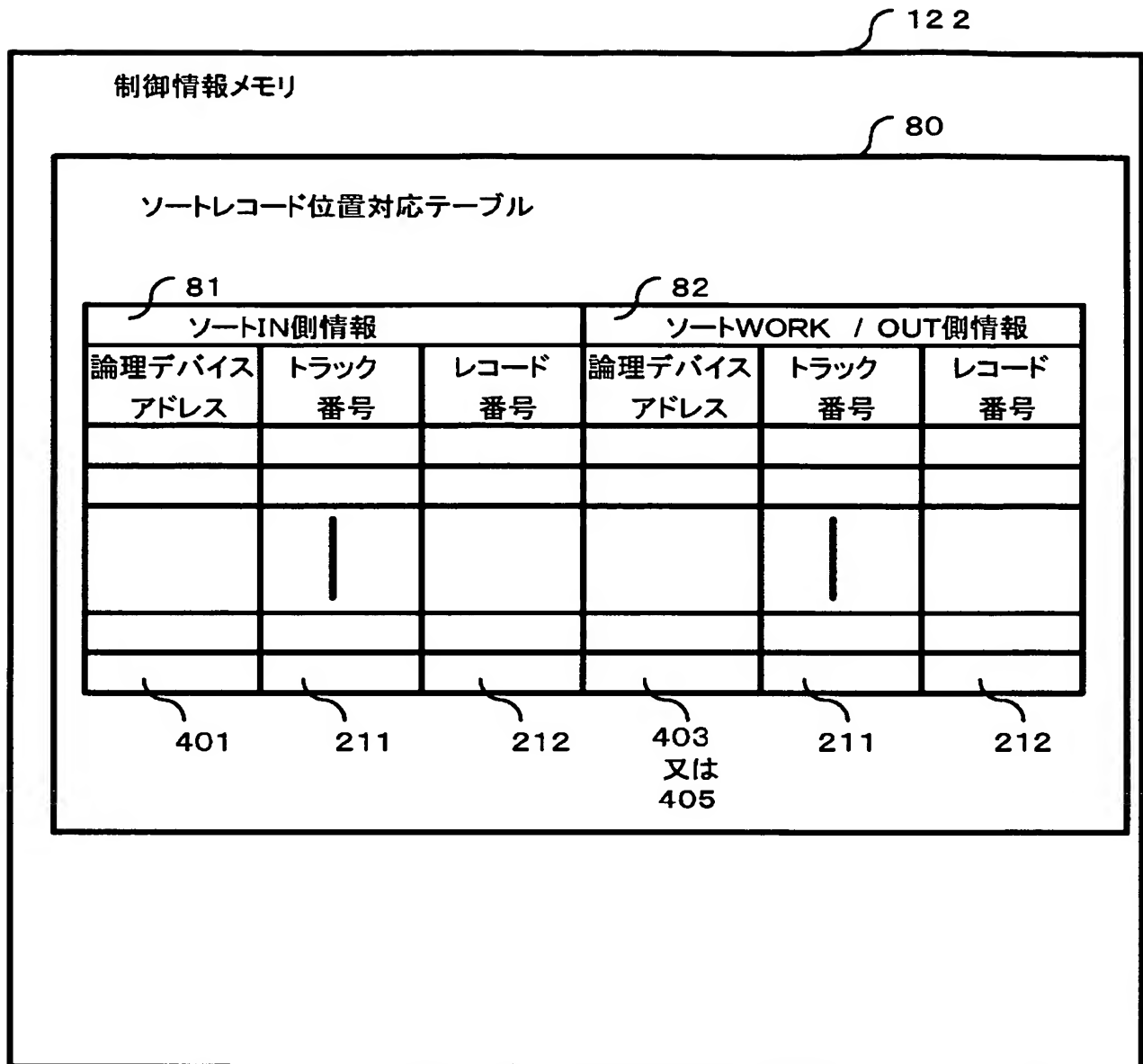


【図 7】



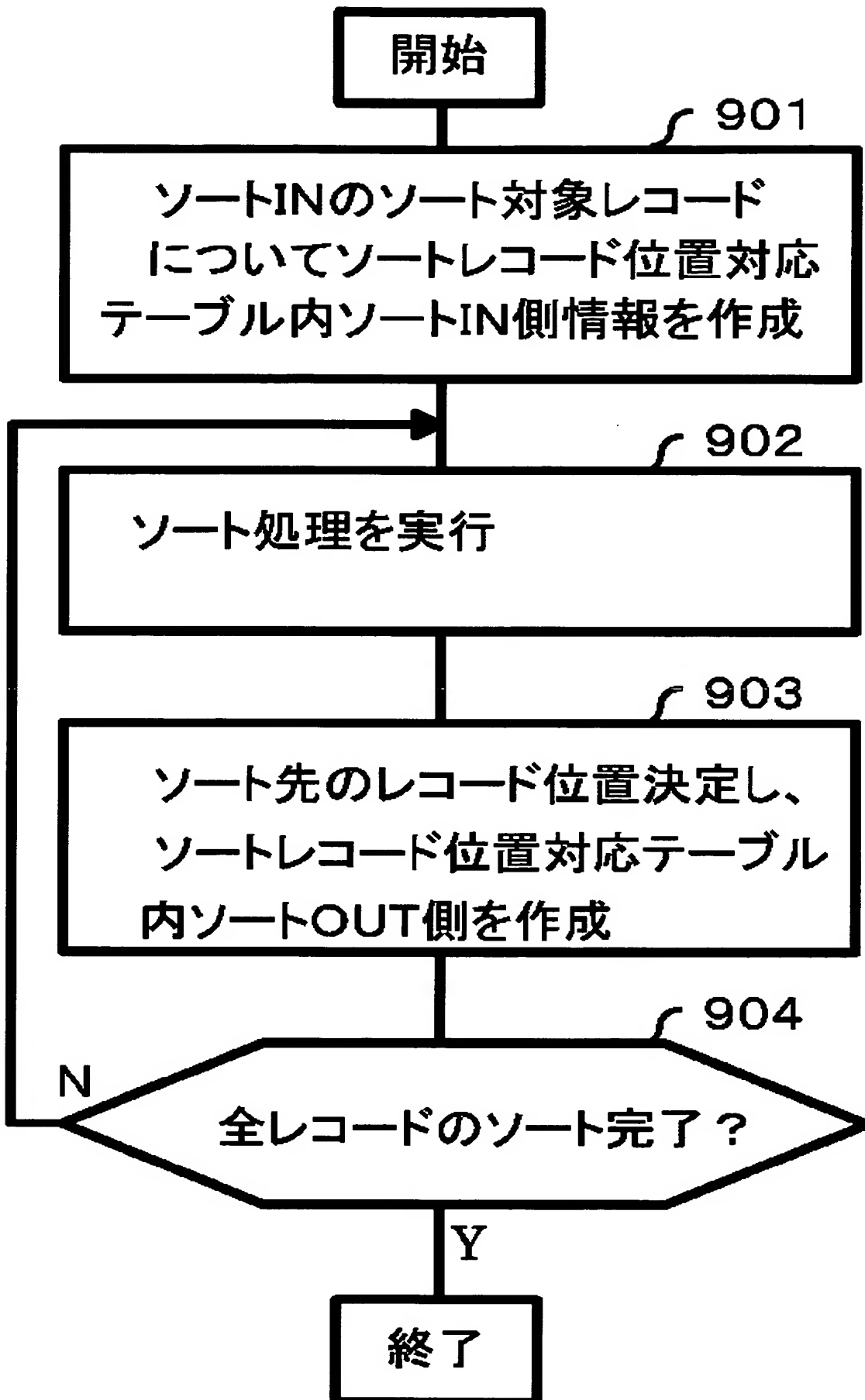


【図 8】



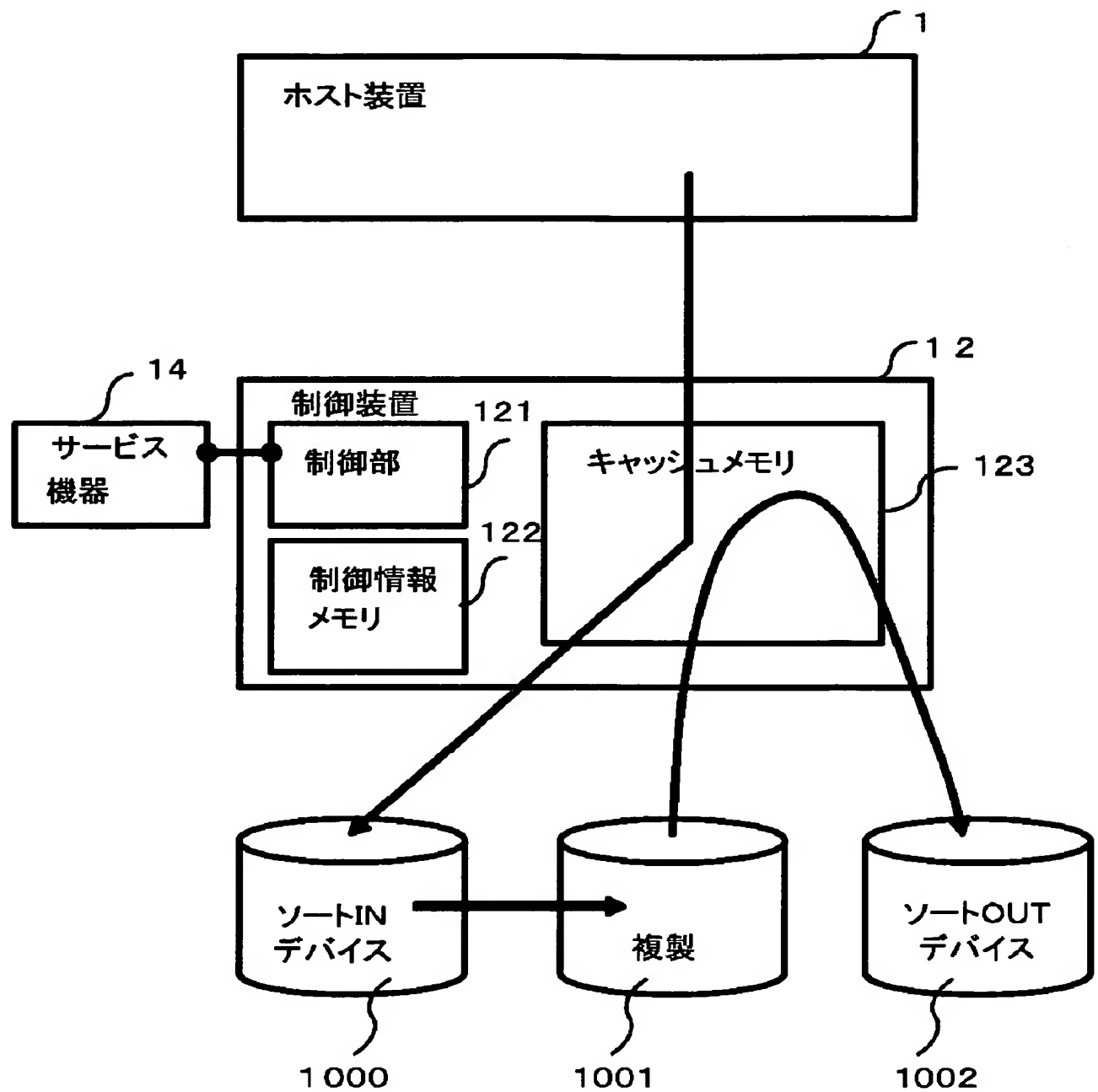


【図 9】



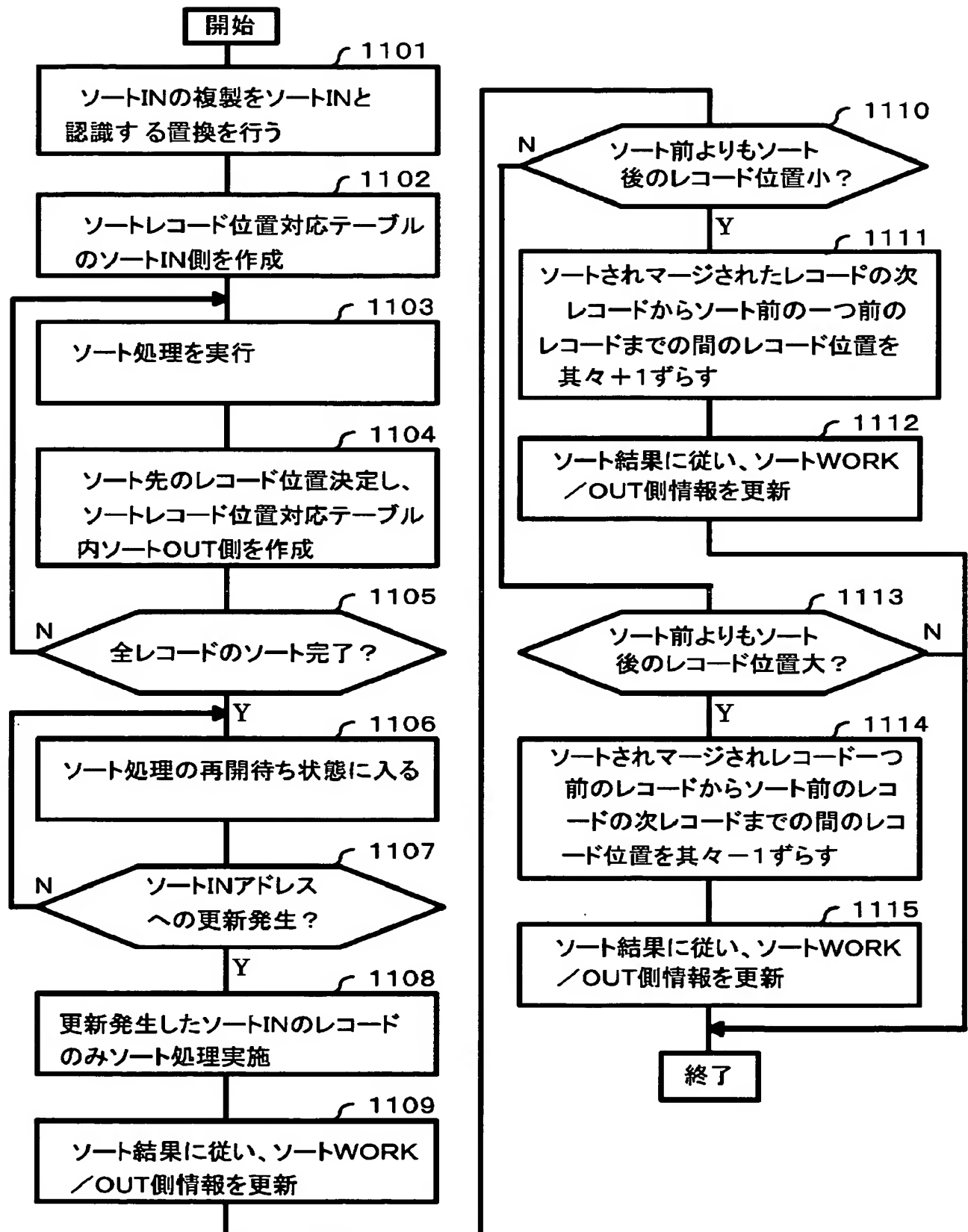


【図10】



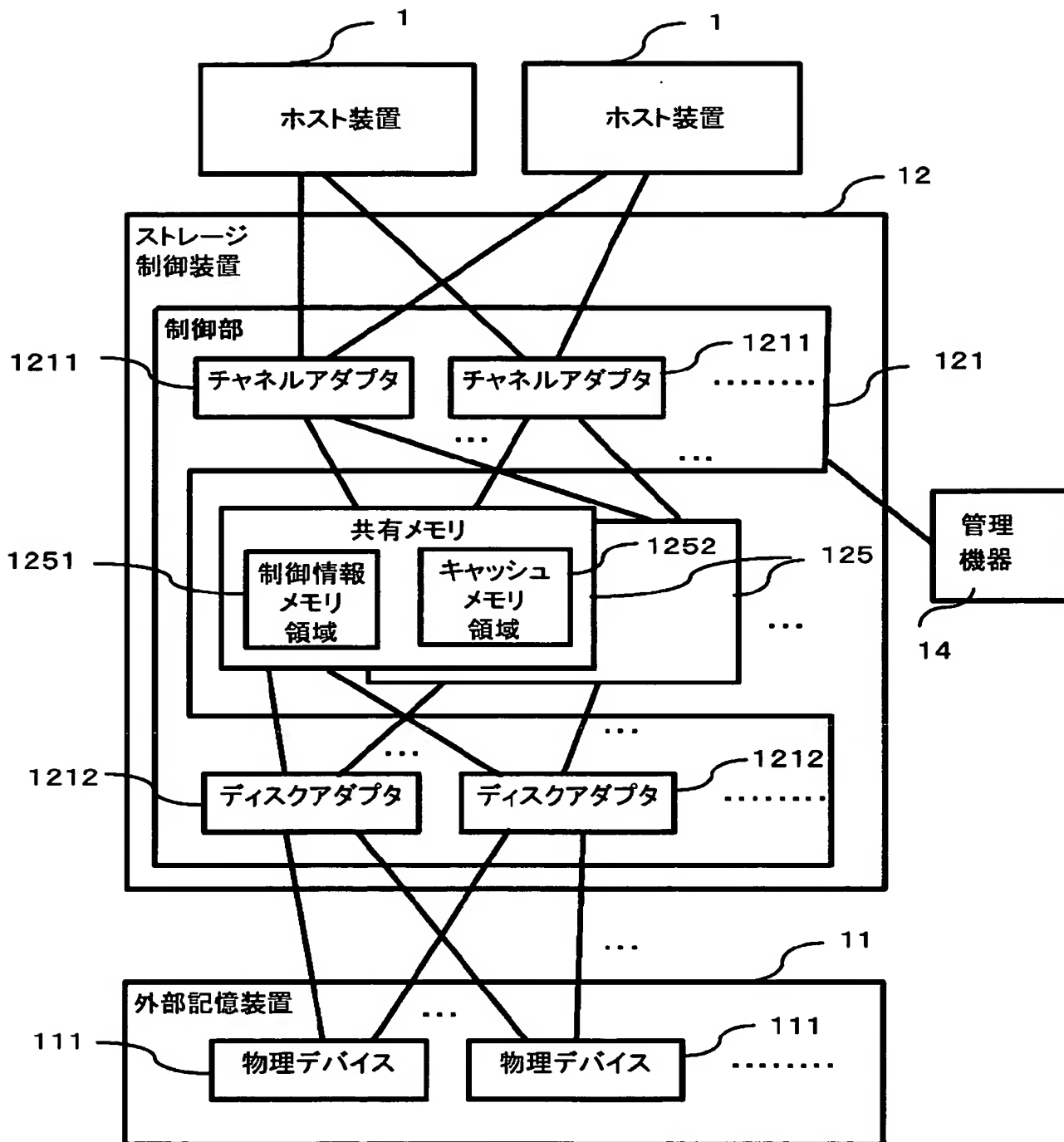


【図 11】





【図 12】





**【書類名】 要約書****【要約】**

**【課題】** ストレージシステム内のデータのソート処理を実行する場合、ホスト装置のオンライン業務への影響を低減する。

**【解決手段】**

ストレージシステム 2 のストレージ制御装置 1 2 が、ホスト装置 1 など外部機器からソート対象レコード群とソートキーを指定したソート指示を受け、外部記憶装置 1 1 内のソート対象コード群をキャッシュメモリ 1 2 3 に読み出し、キャッシュメモリ 1 2 3 と外部記憶装置 1 1 内に確保した作業領域を使って、上記レコード群のソート処理を実施し、ソート後のレコード群を外部記憶装置 1 1 内の所定の出力領域に格納し直す。複数のソート指示を受けた場合、作業領域が重複しないように、自動的に作業領域が割り当てられる。オンライン業務中にレコード更新が発生したとき、更新レコードのスナップショットを取り、スナップショットのソートを行うことで、オンライン業務中にソート処理が行える。

**【選択図】** 図 1 0



認定・付加情報

特許出願の番号	特願 2 0 0 3 - 4 2 8 7 1 8
受付番号	5 0 3 0 2 1 2 6 7 4 9
書類名	特許願
担当官	第七担当上席 0 0 9 6
作成日	平成 1 5 年 1 2 月 2 6 日

< 認定情報・付加情報 >

【提出日】 平成15年12月25日



特願 2 0 0 3 - 4 2 8 7 1 8

出 願 人 履 歴 情 報

識別番号

[ 0 0 0 0 0 5 1 0 8 ]

1. 変更年月日

1 9 9 0 年 8 月 3 1 日

[変更理由]

新規登録

住 所

東京都千代田区神田駿河台 4 丁目 6 番地

氏 名

株式会社日立製作所